



**UCGE Reports
Number 20207**

Department of Geomatics Engineering

**Accuracy Enhancement of Integrated MEMS-IMU/GPS
Systems for Land Vehicular Navigation Applications**

(URL: <http://www.geomatics.ucalgary.ca/links/GradTheses.html>)

by

Walid Abdel-Hamid

January 2005



UNIVERSITY OF
CALGARY

UNIVERSITY OF CALGARY

**Accuracy Enhancement of Integrated MEMS-IMU/GPS Systems for Land
Vehicular Navigation Applications**

by

Walid Abdel-Hamid

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

January, 2005

© Walid Abdel-Hamid 2005

ABSTRACT

This research aims at enhancing the accuracy of land vehicular navigation systems by integrating GPS and Micro-Electro-Mechanical-System (MEMS) based inertial measurement units (IMU). This comprises improving the MEMS-based inertial output signals as well as investigating the limitations of a conventional Kalman Filtering (KF) solution for MEMS-IMU/GPS integration. These limitations are due to two main reasons. The first is that a KF suppresses the effect of inertial sensor noise using GPS-derived position and velocity as updates but within a limited band of frequency. The second reason is that a KF only works well under certain predefined dynamic models and convenient input data that fit these models, which are not attainable with the utilization of MEMS-based inertial technology. Therefore, if the GPS reference solutions are lost, the accuracy of standalone MEMS-IMU navigation will drastically degrade over time.

The Wavelet Multi-Resolution Analysis (WMRA) technique is proposed in this thesis as an efficient pre-filter for MEMS-based inertial sensors outputs. Applying this pre-filtering process successfully improves the sensors' signal-to-noise ratios, removes short-term errors mixed with motion dynamics, and provides more reliable data to the KF-based MEMS-INS/GPS integration module. The results of experimental validation show the effectiveness of the proposed WMRA method in improving the accuracy of KF estimated navigation states particularly position. Moreover, the Adaptive-Neuro-Fuzzy-inference-system (ANFIS)-based algorithm is suggested and assessed to model the variations of the MEMS sensors' performance characteristics with temperature. The

focus is on modeling the gyro thermal variations since it usually dominates the attainable accuracy of INS standalone navigation. Initial results show the efficiency and precision of the proposed ANFIS modeling algorithm. Finally, a new technique augmenting the powerful ANFIS predictor with the traditional KF for improving the integrated MEMS-INS/GPS system performance is presented. The proposed augmentation is utilized either to provide direct corrections to the estimated position by KF during standalone inertial navigation or to supply estimated reference position and velocity error measurements during the absence of GPS solutions, thus keeping the functionality of the KF update engine. Initial test results show the significance of the proposed ANFIS-KF augmentation in reducing position and velocity drifts during GPS outages.

ACKNOWLEDGEMENTS

This work would not have been possible without the help, support, advice and participation of a many people. First, I am indebted to Dr. Naser El-Sheimy, my supervisor, for his patience, dedication and encouragement during my PhD program. I have gained a lot through his exceptional foresight, immeasurable contribution, sincere advice, creative ideas and continuous support. I am truly grateful to him.

I am also indebted to Dr. Gerard Lachapelle, my co-supervisor, for his incessant encouragement and constructive suggestions throughout the course of this research. I would like also to express my gratitude to Dr. Elizabeth Cannon who, through my supervisory committee, provided invaluable feedback to my thesis. I'm really lucky and proud to have such an exemplary supervision committee. I'm also thankful to my examining committee members Dr. Mohamed Ibn Kahla and Dr. Karan Kaler for reading this dissertation and for their feedback.

Special thanks are due to my friend Dr. Husam Kinawi for his sincere efforts in polishing the language of this thesis. Mr. Eun-Hwan Shin, a member of my research group, is acknowledged for providing the INS/GPS KF integration Toolbox that has been used in processing the experimental data in Chapters four and seven. I have also benefited from the stimulating discussions and the support given in many ways by the members of the Department of Geomatics Engineering. In particular, my colleagues in the Mobile-Multi-Sensor-Research-Group: Dr. Xiaoji Niu, Dr. Mohsen Aly, Dr. Kai-Wei Chiang, Mr. Mahmoud El-Gizawy, and Mr. Haying Hou are thanked for their help in the experimental work of this thesis. I deeply appreciate the frank support, significant assistance, and friendship of my close friends Dr. Aboelmagd Nouredin, Dr. Ahmed Osman, Tamer Abdelazeem, Mohamed Elhabiby and Mohamed Abdel-Salam.

This research was supported in part by the research grants from Natural Science and Engineering Research Council of Canada (NSERC) and Geomatics for Informed

Decisions (GEOIDE), Network Centers of Excellence (NCE) to Dr. Naser El-Sheimy, The Egyptian Governmental Scholarship, The Department of Geomatics Engineering Graduate Research Scholarships (GRS) and Special Awards, and the US Institute of Navigation (ION) Student Paper Award.

Through my life, my loving parents were the best teachers I have ever had. You taught me to respect others and helped me build what has brought me this far. You can take all the credit for much of what I have achieved and what I will achieve in the future. My beloved sister, Lamia, has been a constant source of support and inspiration. I am also fortunate to have a loving father and mother-in-law. I owe you all much more than words can express.

Finally, to my sweet and loving wife, Amira, who believed, supported, and stood by me; and to my beloved sons, Mohamed and Omar, to whom I still owe lots of fun hours – you kept me going through it all. I shall never forget your endless love, patience, sacrifice, support and encouragement. Without you, this work would have never seen light. Words will never say how grateful I am.

DEDICATION

To

***My Loving Parents, My Sweet Wife, My Beloved Sons
and My Dear Sister***

“To all of you, I shall be indebted for ever ”

TABLE OF CONTENTS

Approval Page.....	ii
Abstract.....	iii
Acknowledgements.....	v
Dedication.....	vii
Table of Contents.....	viii
List of Tables.....	xi
List of Figures.....	xii
Notation.....	xvi
CHAPTER ONE INTRODUCTION	1
1.1 Background and Problem Statement	2
1.2 Research Objectives	6
1.3 Research Contribution.....	8
1.4 Dissertation Outline.....	9
CHAPTER TWO MEMS INERTIAL SENSORS - PERFORMANCE CHARACTERISTICS AND LIMITATIONS.....	13
2.1 General Characteristics of Inertial Navigation Sensors.....	13
2.2 MEMS Sensors – Types and Operation Principles	19
2.3 Specific Performance Characteristics of MEMS Sensors	27
2.3.1 Output Stability.....	28
2.3.2 Thermal Sensitivity.....	28
2.3.3 Stochastic (Noise) Characteristics	30
2.4 Methods for Analyzing the Noise Characteristic	32
2.4.1 Frequency Analysis of Inertial Signal	33
2.4.2 Time Analysis of Inertial Signals	39
CHAPTER THREE KALMAN FILTERING OF INS/GPS INTEGRATED SYSTEMS.....	46

3.1 GPS and INS Error Characteristics	46
3.1.1 GPS Observation Measurements Errors	47
3.1.2 INS Observation Measurements Errors	54
3.1.3 The Complementary Nature of INS and GPS.....	58
3.2 INS Mechanization.....	59
3.3 INS/GPS Kalman Filtering Algorithm	61
3.3.1 Dynamic and Measurement Models	61
3.3.2 Kalman Prediction and Update Equations.....	64
3.4 INS/GPS Integration Schemes	66
3.5 Limitations of INS/GPS Kalman Integration	72
CHAPTER FOUR WAVELT MULTI-RESOLUTION ANALYSIS OF MEMS INERTIAL SIGNAL	74
4.1 Observability of Different Inertial Error Bandwidths	74
4.2 Power Spectrum In Static and Kinematic Modes.....	78
4.3 Wavelet Multi-Resolution Analysis and Thresholding.....	82
4.3.1 Continuous and Discrete Wavelet Transform.....	83
4.3.2 Wavelet Multi-Resolution Analysis (WMRA).....	87
4.3.3 Wavelet Thresholding.....	91
4.4 Wavelet Decomposition of MEMS Inertial Output Signal	94
4.5 Experimental Validation of The Proposed WMRA	96
CHAPTER FIVE FUZZY LOGIC BASED PREDICTION TECHNIQUES	104
5.1 Concepts of Fuzzy Logic.....	105
5.2 Basic Elements of a Fuzzy Inference System (FIS)	109
5.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)	113
5.3.1 Consequent Parameters Estimation	118
5.3.2 Antecedent Parameters Tuning.....	120
5.3.3 Limitation of ANFIS.....	123
5.4 Principles of Fuzzy OLS	124
5.4.1 TS Model Identification Through Optimal Clustering	126
5.4.2 Consequent and Antecedent Parameters Selection.....	130
5.5 Summary.....	132

CHAPTER SIX	THERMAL CHARACTERISTICS OF MEMS INERTIAL SENSORS	134
6.1	Thermal Characteristics and Variation of MEMS Inertial Sensors Output.....	134
6.2	Thermal Testing Approaches	140
6.3	Results of Lab Thermal Testing of Various MEMS Inertial Sensors	141
6.3.1	Test Description.....	143
6.3.2	Thermal Variation of Sensors Calibration Parameters	145
6.4	ANFIS Prediction of Thermal Drift	151
6.4.1	Structure of the Implemented FIS.....	151
6.4.2	Results of Applying The Proposed Technique	153
6.4.3	Numerical Analysis.....	158
CHAPTER SEVEN	FUZZY PREDICTION OF MEMS-BASED INS/DGPS NAVIGATION ERROR STATES	161
7.1	An Augmented Fuzzy-Kalman For Position Error Prediction	161
7.1.1	Methodology.....	163
7.1.2	Experimental Validation.....	168
7.1.3	Results and Analysis.....	170
7.2	An Augmented Fuzzy-Kalman to Preserve Update Engine Functionality.....	175
7.2.1	Methodology.....	176
7.2.2	Experimental Validation.....	179
7.2.3	Results and Analysis.....	180
CHAPTER EIGHT	SUMMARY, CONCLUSIONS AND RECOMMENDATIONS	191
8.1	Summary.....	191
8.2	Conclusions and Recommendations.....	194
REFERENCES	200

LIST OF TABLES

Table 1.1	Performance of Different grades of inertial sensors [Shin, 2001].....	8
Table 2.1	Classification of MEMS sensors according to mass position detection	22
Table 2.2	Classification of MEMS sensors according to fabrication processes.....	23
Table 3.1	Decentralized against centralized INS/GPS integration.....	71
Table 4.1	Positioning accuracy enhancement after WMRA	103
Table 6.1	Technical specifications of the AHRS400CA MEMS-based IMU.....	142
Table 6.2	Technical specifications of the MotionPakII MEMS-based IMU	143
Table 7.1	The specifications of the ISI IMU.....	169
Table 7.2	Positioning error reduction after applying the augmented solution	173
Table 7.3	Positioning accuracy enhancement after applying the augmented solution	174
Table 7.4	Maximum error values and RMSE of position and velocity during outage (3).....	188
Table 7.5	Maximum error values and RMSE of position and velocity during outage (5).....	188
Table 7.6	Maximum error values and RMSE of position and velocity during outage (6).....	189
Table 7.7	The proposed Fuzzy-Kalman versus the traditional KF.....	190

LIST OF FIGURES

Figure 2.1	Typical slopes in $\log_{10} - \log_{10}$ of one-sided PSD versus frequency (accelerometer acceleration data or gyroscope angle rate data).	37
Figure 2.2	Ensemble and frequency average of the one-sided PSD for the MEMS gyro CRS03	38
Figure 2.3	Ensemble and frequency average of the one-sided PSD for the MEMS gyro CRS02	38
Figure 2.4	Autocorrelation function of a 1 st order Gauss-Markov random process and its correlation time.....	41
Figure 2.5	Autocorrelation functions of the Cross-Bow AHRS300CA Z-gyro at three different sensor temperatures [-30°, 0° and +30°C]	41
Figure 2.6	Power spectral density and autocorrelation function of white noise	42
Figure 2.7	Power spectral density and autocorrelation function of first order Gauss-Markov Random Process.....	43
Figure 2.8	Autocorrelation function of a second order Gauss-Markov Random Process.....	44
Figure 2.9	Power spectral density and autocorrelation function of infinite order Gauss-Markov Random Process (bias)	45
Figure 3.1	Functional block diagram of KF-based IMU/GPS Integration during GPS signal availability	66
Figure 3.2	Architecture of a centralized INS/GPS integration (open loop) – after [El-Sheimy, 2004]	69
Figure 3.3	Architecture of a centralized INS/GPS integration (closed loop) – after [El-Sheimy, 2004]	70
Figure 3.4	Architecture of a decentralized INS/GPS integration (open loop) - after [El-Sheimy, 2004]	70
Figure 3.5	Architecture of a decentralized INS/GPS integration (closed loop) – after [El-Sheimy, 2004]	71
Figure 3.6	Functional block diagram of KF-based IMU/GPS Integration during GPS signal availability	73
Figure 4.1	Error bandwidth in an INS system: a) raw measurements, b) after standard processing (F_S is the sampling frequency) [Skaloud et al., 1999]	77

Figure 4.2	Power spectrum of MEMS lateral acceleration in static mode.....	79
Figure 4.3	Power spectrum of MEMS lateral acceleration in kinematic mode	80
Figure 4.4	Power spectrum of MEMS pitch rate in static mode	80
Figure 4.5	Power spectrum of MEMS pitch rate in kinematic mode.....	81
Figure 4.6	A scalogram illustrating Short-time Fourier Transform [Robertson et al., 1996].....	83
Figure 4.7	Example of wavelet functions.....	86
Figure 4.8	A scalogram representing Discrete Wavelet Transform [Robertson et al., 1996].....	87
Figure 4.9	Wavelet multi-resolution analysis considering three levels of decomposition.....	91
Figure 4.10	Two different threshold operators (a) hard threshold (b) soft threshold	93
Figure 4.11	MEMS forward acceleration signal before and after wavelet thresholding.....	95
Figure 4.12	MEMS pitch rate signal before and after wavelet thresholding	96
Figure 4.13	Test field trajectory with the simulated GPS outage periods indicated.....	97
Figure 4.14	Position accuracy during GPS outage (1) – Raw versus de-noised INS measurements [denoising with soft threshold WMRA]	98
Figure 4.15	Position accuracy during GPS outage (1) – Raw versus de-noised INS measurements [denoising with hard threshold WMRA].....	99
Figure 4.16	Position accuracy during GPS outage (2) – Raw versus de-noised INS measurements [denoising with soft threshold WMRA]	99
Figure 4.17	Position accuracy during GPS outage (2) – Raw versus de-noised INS measurements [denoising with hard threshold WMRA]	100
Figure 4.18	Position accuracy during GPS outage (3) – Raw versus de-noised INS measurements [denoising with soft threshold WMRA]	100
Figure 4.19	Position accuracy during GPS outage (4) – Raw versus de-noised INS measurements [denoising with soft threshold WMRA]	101
Figure 5.1	Basic components of a Fuzzy Inference System (FIS).....	109
Figure 5.2.	A simple TS-FIS structure of two inputs, two rules and a single output	114

Figure 5.3.	ANFIS learning using hybrid technique	117
Figure 6.1	Extracting the thermal trend of the Cross-Bow AHRS300CA Z-axis accelerometer using wavelet decomposition	136
Figure 6.2	Extracting the thermal trend of the Cross-Bow AHRS300CA Z-axis gyroscope using wavelet decomposition	136
Figure 6.3	Cross-Bow AHRS400CA MEMS-based IMU	141
Figure 6.4	Systron Donner's MotionPakII MEMS-based IMU	142
Figure 6.5	SK-2101 Temperature chamber	144
Figure 6.6	Thermal variation of the Crossbow AHRS400CA X & Y-axis accelerometer biases	147
Figure 6.7	Thermal variation of the Crossbow AHRS400CA X & Y-axis accelerometer scale factors	147
Figure 6.8	Thermal variation of Systron Donner's MPakII X & Y-axis accelerometer biases	148
Figure 6.9	Thermal variation of Systron Donner's MPakII X & Y-axis accelerometer scale factors	148
Figure 6.10	Thermal variation of the Crossbow AHRS400CA X & Y-axis gyroscope biases	149
Figure 6.11	Thermal variation of the Systron Donner's MPakII X & Y-axis gyroscope biases	149
Figure 6.12	Single inputs/single output FIS structure for ANFIS learning.....	152
Figure 6.13	RMS error of ANFIS prediction and step size during the training process (Crossbow X-gyro)	155
Figure 6.14	Prediction error of developed ANFIS algorithm when applied on test data (Crossbow X-gyro)	155
Figure 6.15	Initial and final tuned membership functions of Crossbow X-gyro thermal trend algorithm.....	156
Figure 6.16	RMS error of ANFIS prediction and step size during training process (MPakII X-gyro)	156
Figure 6.17	Prediction error of developed ANFIS algorithm when applied on test data (MPakII X-gyro).....	157
Figure 6.18	Initial and Final tuned membership functions of MPakII X-gyro thermal trend algorithm.....	157
Figure 7.1	Extracting the input/output data entries for ANFIS learning.....	165

Figure 7.2	Corrections provided by ANFIS to the KF predicted position errors	165
Figure 7.3	Working mechanism of the augmented ANFI-KF system for providing direct position corrections (learning mode).....	166
Figure 7.4	Working mechanism of the augmented ANFI-KF system for providing direct position corrections (prediction mode)	166
Figure 7.5	Test field trajectory with simulated GPS outages indicated	170
Figure 7.6	Position accuracy during GPS outage (2) before and after ANFIS position corrections.....	171
Figure 7.7	Position accuracy during GPS outage (4) before and after ANFIS position corrections.....	171
Figure 7.8	Position accuracy during GPS outage (7) before and after ANFIS position corrections.....	172
Figure 7.9	Input/output data extraction during ANFIS training.....	177
Figure 7.10	Providing simulated reference error measurements by ANFIS	178
Figure 7.11	Working mechanism of the augmented ANFI-KF system for preserving KF update engine functionality (prediction mode).....	179
Figure 7.12a	Kalman estimated velocity drift during outage (3).....	182
Figure 7.12b	ANFIS-KF estimated velocity drift during outage (3).....	182
Figure 7.13a	Kalman estimated velocity drift during outage (5).....	183
Figure 7.13b	ANFIS-KF estimated velocity drift during outage (5).....	183
Figure 7.14a	Kalman estimated velocity drift during outage (6).....	184
Figure 7.14b	ANFIS-KF estimated velocity drift during outage (6).....	184
Figure 7.15a	Kalman estimated position drift during outage (3).....	185
Figure 7.15b	ANFIS-KF estimated position drift during outage (3).....	185
Figure 7.16a	Kalman estimated position drift during outage (5).....	186
Figure 7.16b	ANFIS-KF estimated position drift during outage (5).....	186
Figure 7.17a	Kalman estimated position drift during outage (6).....	187
Figure 7.17b	ANFIS-KF estimated position drift during outage (6).....	187

NOTATION

1. Conventions

- 1.1 Vectors, matrices and functions are typed in boldface.
- 1.2 Vectors are represented by lower-case letters.
- 1.3 Matrices are represented by upper-case letters.
- 1.4 Functions are represented by either upper-case or lower-case letters.
- 1.5 A dot above a vector; a matrix or a quantity indicates a time differentiation.
- 1.6 A “vector” is always considered as three-dimensional. A superscript indicates the particular coordinate frame in which the vector is represented. For example:

$\mathbf{r}^i = (r_x^i, r_y^i, r_z^i)^T$ represents the components of the vector \mathbf{v} in the i -frame.

- 1.7 Rotation (transformation) matrices between two coordinate frames are denoted by \mathbf{R} . The two coordinate frames are indicated by a superscript and a subscript. For example:

\mathbf{R}_i^j represents a transformation matrix from the i -frame to the j -frame.

- 1.8 Angular velocity between two coordinate frames represented in a specific coordinate frame can be expressed either by a vector $\boldsymbol{\omega}$ or by the corresponding skew-symmetric matrix $\boldsymbol{\Omega}$. A superscript and two subscripts will be used to indicate the corresponding coordinate frames. For example:

$\boldsymbol{\omega}_{ij}^k = (\omega_x, \omega_y, \omega_z)^T$ or $\boldsymbol{\Omega}_{ij}^k = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}$ describes the angular

velocity between the i -frame and the j -frame represented in the k -frame.

2. Coordinate Frames

2.1 Operational Inertial Frame (i-frame)

- Origin: at the center of mass of the Earth.
- X-axis: towards the mean vernal equinox.
- Y-axis: completes a right-handed system.
- Z-axis: towards the north celestial pole.

2.2 Earth-Fixed Frame (e-frame)

- Origin: at the center of mass of the Earth.
- X-axis: towards the mean Greenwich meridian in the equatorial plane.
- Y-axis: completes a right-handed system.
- Z-axis: direction of mean spin axis of the Earth.

The rotation of the e-frame with respect to the i-frame is expressed in terms of a constant Earth rotation rate ω_e such that:

$$\omega_{ei}^e = (0, 0, \omega_e)^T = (0, 0, 15.04106853 \text{ deg/h})^T$$

2.3 Body Frame (b-frame)

- Origin: at the center of the orthogonal IMU accelerometer triad.
- X-axis: towards the right side of the IMU (when viewed from the front IMU panel).
- Y-axis: towards the direction opposite to the IMU front panel.
- Z-axis: pointing up and completing a right-handed system.

2.4 Local-Level Frame (l-frame)

- Origin: at the center of the b-frame.
- X-axis: towards the ellipsoidal east.
- Y-axis: towards the ellipsoidal north.
- Z-axis: pointing outward along the ellipsoidal normal.

3. List of Acronyms

$\Delta\nabla$	Double differencing operator
ANFIS	Adaptive-Neuro-Fuzzy-Inference-System
AO	Alternating optimization
AR	Autoregressive
CG	Centre of area or gravity
CWT	Continuous Wavelet transform
DB	Database
DFT	Discrete Fourier transform
DGPS	Differential GPS
DWT	Discrete Wavelet transform
FIS	Fuzzy-Inference-System
FT	Fourier transform
GPS	Global Positioning System
gyro	Gyroscope
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
KB	Knowledge base
KF	Kalman Filtering
LSE	Least-squares estimate
max	Maximum
MEMS	Micro-Electro-Mechanical-Systems
MF	Membership function

min	Minimum
MISO	Multi-input/multi-output
MV	Maximum value
NAV	Navigation
OLS	Orthogonal Least Squares
POS	Position
ppm	Part per million
PSD	Power spectral density function
PVA	Position, velocity and attitude
RB	Rule base
RMSE	Root-Mean-Square-Error
SPP	Single point positioning
STFT	Short time Fourier transform
TS	Takagi-Sugeno
WMRA	Wavelet Multi-resolution Analysis
WT	Wavelet transform

4. List of Symbols

g	Gravitational force
α	Thermal expansion coefficient
τ	Correlation time
E	Expectation of a process
T	Temperature
Hz	Hertz
ω_o	Cut-off frequency
Φ_{xx}	Power spectral density of a random process x
φ_{xx}	Autocorrelation function of a random process x
$\delta(\tau)$	Dirac function
σ^2	Variance
σ	Standard deviation
$\delta\rho$	Error in the measured range between satellite and receiver
K	Kalman gain matrix
K_b	Boltzmann's constant
$\psi(t)$	Wavelet basis function
$\delta\phi$	Error in latitude
$\delta\lambda$	Error in longitude
δh	Error in altitude (height)
F	Fuzzy set

U	Universe of discourse
ε	ANFIS learning rate
κ	Training step size
μ_F	Membership function
$^{\circ}\text{C}$	Celsius degrees
dv	Velocity error
dp	the position error
f^b	3x1 vector of linear accelerations as measured in the IMU body frame
ω_{ib}^b	3x1 vector of angular velocities of the IMU body frame with respect to inertial frame as seen from the IMU body frame

CHAPTER ONE

INTRODUCTION

During the last decade, low-cost Micro-Electro-Mechanical-System (MEMS) based inertial sensors have been steadily emerging into the market place. Their miniature design and lightweight result in many advantages such as low power consumption, wide bandwidth, survivability and cost reduction [Stephen, 2000]. These are basic requirements in most commercial applications particularly the automobile market. These attractive advantages have led to remarkable research progress in the field of MEMS inertial sensors under strong support from both government and industries [Varadan et al., 2001].

On the negative side, the lightweight and miniature design of MEMS sensors have led to some critical disadvantages and limitations like low sensitivity to small dynamics which can be part of the motion of interest [Davis, 1998]. Limitations and shortcomings of MEMS sensors arise mainly from the relatively large, random and rapidly varying errors accompanying their measured output. These errors are consequently difficult to model through traditional methods used with other non-MEMS sensors.

The goal of this chapter is henceforth to present background information on the evolution of MEMS technology and its increased adoption in the inertial navigation field. The main advantages of MEMS inertial sensors as well as the basic limitations and shortcomings,

that significantly degrade their performance, are presented. The problem of optimizing the estimation procedure of the MEMS-INS/GPS integrated system is stated. The general approaches that can be followed to enhance MEMS sensors performance are proposed in order to improve the overall performance of integrated navigation systems utilizing these sensors. Finally, research objectives, research contribution, and the dissertation outline are given.

1.1 Background and Problem Statement

In addition to the miniature size of MEMS sensors, the fabrication processes of MEMS sensors make them very sensitive to the changes in the surrounding environmental conditions like temperature, pressure, electric and magnetic fields, etc. These changes cause the output of MEMS sensors to vary rapidly, widely and sometimes randomly which make it difficult to appropriately model such variation [Nebot and Durrant-Whyte, 1999]. This high sensitivity to the surrounding environmental conditions adds more error types and possibly, higher errors than those of traditional non-MEMS sensors. Unless those errors are categorized, modelled and determined utilizing special algorithms or techniques, the overall performance of MEMS sensors will be degraded when used in different applications, particularly inertial navigation. Henceforth, we first provide a categorization of the different errors of MEMS-based inertial sensor management, and second, we study the impact of such categorization in specific navigation applications to highlight the problems addressed by this thesis.

Errors of MEMS-based inertial sensor measurements can be classified into two categories based on their ability to be determined or modelled:

- First, deterministic errors due to large bias and drift values and their relatively wide variation with surrounding environmental conditions, particularly temperature. The variation of the sensor bias can take place either during the same run, known as in-run bias instability, or from turn on to another namely run-to-run or turn-on-to-turn-on bias instability. Part of the in-run sensor bias instability, most probably caused by the variation of environmental conditions, is also deterministic and can be quantified or modeled based on intensive and careful lab testing [Abdel-Hamid, 2004]. Nevertheless, the remaining part is hard to model and thus is considered as a random or stochastic part, which belongs to the second error category.
- Second, stochastic errors due to the relatively high noise level accompanying the output measurements of MEMS sensors. This output noise introduces short-term errors with frequency ranges coinciding with that of the motion of interest beside those caused by secondary dynamics and engine vibrations in kinematic mode [Abdel-Hamid et al., 2004]. Stochastic errors also arise from the random part of the in-run instability of sensor biases as well as the run-to-run bias instability.

The last two decades have seen an increasing trend in the use of positioning and navigation (POS/NAV) technologies in land vehicle (mainly for automobile navigation) applications. Two categories of navigation systems are presently utilized for vehicle positioning. These are the Global Positioning System (GPS) and inertial navigation system (INS) technology [Schwarz and El-Sheimy, 1999]. In fact, both systems can be

integrated together to provide reliable navigation and positioning solutions for moving vehicles [Hide et al., 2002]. The integration of the two systems overcomes each of their shortcomings and offers several advantages. In other words, the integration process results in a system that has superior performance in comparison with either a GPS or an INS stand-alone system. The superior performance of this integration is mainly due to the complementary performance characteristics and contrary error behaviour of GPS and INS systems. For instance, GPS derived positions have good long-term stability and approximately white noise characteristics over the whole frequency range. On the other hand, short-term navigation errors from the INS are relatively small and less noisy, but they degrade rapidly without bound over time [Mohinder and Angus, 2001]. The GPS-derived positions and velocities are therefore excellent external measurements for updating the INS with position parameters, thus improving its long-term accuracy. Similarly, the INS can provide precise position and velocity data, for GPS signal acquisition and reacquisition after outages, as well as the necessary orientation parameters for the moving platform. In general, the fact that redundant measurements are available for the determination of the vehicle trajectory parameters greatly enhances the reliability of the system [Abdelazim et al., 2004].

In order to optimally combine both systems, an applied optimal estimation tool like Kalman filtering (KF) is used to fuse INS and GPS data in real-time. Based on an error model of the different navigation system parameters (including stochastic models for sensor errors), a KF solution may be capable of providing a reliable estimate of the position, velocity and attitude components of the moving platform [El-Sheimy, 2004]. In

addition, the KF may predict the vehicle position during GPS signal blockage relying on the localized INS outputs and the pre-designed error model.

In the decentralized architecture of a KF based GPS/INS integration, two filters are implemented and performed independently; one for GPS and the other for INS. The INS filter is the primary one that uses the raw sensor data of INS to estimate the state (position, velocity and attitude) along the trajectory through a process called INS mechanization [Titterton & Weston, 1997]. The output of the GPS filter is then used to update the INS filter to help with the estimation of the INS long-term errors. In order to estimate the INS errors, the KF depends on measurements and a dynamics model [Bruton et al., 1999]. The quality of the final estimates of the state depends therefore on the quality of both the measurements being made and the models being implemented [El-Sheimy, 2004].

However, the present KF-based INS/GPS integration techniques may not be optimal for MEMS-based navigation systems. This is due to two main reasons. The first is that a KF suppresses the effect of inertial sensors noise using GPS position and velocity update but within a limited band of frequency [Bruton et al., 1999]. If the same technique is applied for MEMS-based INS, the overall accuracy may be jeopardized due to the significance of the MEMS-based inertial sensor errors that are mixed with the motion dynamics [Skaloud et al., 1999]. The second reason is that a KF only works well under certain predefined dynamic models and convenient input data that fit these models. Nevertheless, it is extremely difficult to set a certain stochastic model for each MEMS-based inertial sensor

(gyroscopes and accelerometers) that works efficiently in all environments and reflects the long-term behaviour of these sensor errors [Hide et al., 2002].

Therefore, if a KF is implemented for MEMS-based INS integration with GPS, the overall accuracy may be jeopardized due to the significance of those MEMS-based inertial sensor errors that are mixed with the motion dynamics. As a result, if the GPS reference measurements are blocked or attenuated the accuracy of the KF predicted navigation states, in MEMS-INS standalone mode, will drastically degrade over time [Abdel-Hamid et al., 2004]. To overcome this problem, it is essential to reduce the noise level at the output of MEMS-based inertial sensors and separate the sensor errors from motion dynamics prior to processing their measurements by the KF-based INS/GPS module. Moreover, another efficient prediction tool or algorithm has to augment the KF in order to replace or at least enhance the poor KF prediction performance during the absence of update or reference GPS measurements. This algorithm should be highly capable of handling and representing non-linear and imprecise systems.

1.2 Research Objectives

The overall objectives of this thesis are as follows: The first is to improve the performance of MEMS-based inertial sensors, for integration with GPS for land vehicle navigation applications, so that it becomes close or comparable to that of tactical grade sensors. Table 1.1 lists the main performance characteristics of the different grades of inertial sensors. The second is to enhance the performance of the conventional KF

solution of MEMS-INS/GPS integration particularly during GPS signal or solution outage. Under the overall objective, the following tasks are addressed:

1. Investigate the general performance characteristics of MEMS-based inertial sensors and their variations with the change of environmental conditions, particularly temperature. This investigation will help find a proper technique or algorithm to efficiently quantify or model the variation of those performance characteristics.
2. Analyze the spectral and temporal behavior of MEMS sensor output signals to eliminate long term and short-term errors contaminating the true signals. A part of these errors is not detectable by GPS update measurements utilized in a KF and hence should be pre-filtered through other powerful analyzing tools. A cascade-denoising algorithm was proposed by Chiang [2004] where he achieved an enhancement of 30% percentage in the performance of KF over short-term for low-cost INS in standalone navigations. Alternatively, Wavelet Multi-Resolution Analysis (WMRA) is adopted here, for its simplicity and efficiency, to decompose the MEMS inertial signal and separate the true motion signal from other interfering signals.
3. Develop an efficient predictor that can compensate for the loss of reference measurements, vital for KF, during GPS outages. Chiang [2004] has successfully applied an intensively trained Neural-networks prediction algorithm. He utilized the difference between GPS measurements, regardless of their efficiency, and KF estimates of navigation states (position and velocity) as the desired outputs for training his proposed Neural network. In this thesis, Fuzzy logic-based prediction

networks are employed, instead, for their capability to treat highly imprecise and nonlinear problems like the drift characteristics of MEMS sensor output signals. In addition, the desired outputs for Fuzzy training are determined as the difference between KF predicted and updated navigation states at each GPS measurements' epoch while the GPS solution is available. In this way, the efficiency (accuracy) of the GPS update measurements is considered during the training procedure.

Table 1.1 Performance of Different grades of inertial sensors [After El-Sheimy, 2004, and Shin, 2001]

Grade	Navigation		Tactical		Automotive		Consumer	
	Gyro	Accel.	Gyro	Accel.	Gyro	Accel.	Gyro	Accel.
Bias	0.005-0.01 (deg/hr)	5 – 10 (μg)	1-10 (deg/hr)	200-500 (μg)	150-180 (deg/hr)	1200 (μg)	360 (deg/hr)	2400 (μg)
Scale Factor	5 – 50 ppm	10-20 ppm	200-500 ppm	400-1000 ppm	-----	-----	-----	-----
Noise	.002-.005 deg/hr/ $\sqrt{\text{Hz}}$	5 – 10 μg /hr/ $\sqrt{\text{Hz}}$	0.2-0.5 deg/hr/ $\sqrt{\text{Hz}}$	200-400 μg /hr/ $\sqrt{\text{Hz}}$	-----	-----	-----	-----
Positioning Error	≈ 2 (km/hr)		20-40 (km/hr)		≈ 2 (km/min)		≈ 3 (km/min)	

1.3 Research Contribution

In this dissertation, the WMRA is adopted as an alternative to the traditionally used low-pass and high-pass filters for effective multi-level decomposition of MEMS inertial signals. The purpose of this decomposition is to improve the quality of MEMS sensors raw measurements through denoising as well as separating the true motion from

unwanted secondary dynamics and short-term errors before integrating the inertial measurements with GPS in a KF solution. A thorough testing and analysis is carried out using real kinematic navigation data to show the effectiveness and suitability of the WMRA technique. It is shown that the WMRA of raw inertial measurements significantly improves the accuracy of the KF final estimates of navigation states (particularly position) during GPS outages.

The adaptive-neuro-fuzzy-inference-system (ANFIS) based prediction network is also presented in this thesis as a tool to handle imprecise and highly non-linear behaviours of MEMS sensor output. Unlike WMRA, which works at the sensor raw measurements level, ANFIS operates directly at the final navigation output level. ANFIS has been adopted for two different applications; First, to model the thermal variation of MEMS sensor performance characteristics, particularly bias and scale factor. Second, to be augmented with the conventional KF for MEMS-INS/GPS integration in order to handle the problem of accuracy degradation in INS standalone mode. Experimental validation is carried out using various data from lab and real navigation tests. The proposed ANFIS is shown to be effective to successfully handle the problem in both situations, and improves the overall accuracy of the integrated MEMS-INS/GPS system with a significant percentage of 70 to 90%.

1.4 Dissertation Outline

In Chapter Two, a general overview of the MEMS-based inertial sensors (their types, fabrication process, operation principles and specific performance characteristics) is

given. Methods of analyzing MEMS sensors output noise including time and frequency analysis are also discussed.

Chapter Three covers the KF theory and implementation for the INS/GPS integrated system. The problem of the integrated system is discussed and its KF formulation is outlined. Details of the INS/GPS KF structure are also discussed. The basic schemes commonly used for integrating inertial and GPS are given. Limitations and shortcomings of the KF integration of INS/GPS are listed.

The WMRA is analysed and verified in Chapter Four as a proposed tool to improve the performance of MEMS-based inertial sensors. First, an analysis of MEMS-based accelerometer and gyroscope measurements in the frequency domain is offered. WMRA is then implemented to improve the sensors signal to noise ratio, remove sensor errors mixed with motion dynamics and provides more reliable data to the KF based MEMS-INS/GPS integration module. Results of experimental validation of WMRA are illustrated and analyzed.

In Chapter Five, the fundamentals and theoretical basis of two proposed Fuzzy-based techniques, to improve the integrated MEMS-INS/GPS system performance, are presented. The first proposed fuzzy-based technique mainly adopts the Adaptive Neuro-Fuzzy Inference System (ANFIS) while the other utilizes the so-called Fuzzy Orthogonal Least-Squares (Fuzzy-OLS) algorithm. The two proposed Fuzzy techniques are mainly

used to predict the position and velocity errors during GPS signal outages where the conventional Kalman algorithm is working only in prediction mode.

In Chapter Six, the effect of thermal variation on MEMS-based inertial sensors is discussed first; then the results of lab thermal testing of two MEMS-based IMUs (Crossbow AHRS400CA and MotionPak II) are presented. Finally, a proposed modeling of MEMS inertial output variation, based on the ANFIS, is illustrated and experimentally validated.

Performance improvement of integrated inertial measurement units (IMUs), utilizing MEMS sensors, and GPS is described in Chapter Seven. Two techniques augmenting a KF and Fuzzy principles, for improving the integrated system performance, are presented. The first proposed technique provides direct corrections to the estimated position by a KF during standalone inertial navigation mode. The second proposed technique supplies simulated reference position and velocity error measurements to the KF update engine keeping its functionality during GPS signal outages. Both proposed techniques have been experimentally verified.

Finally, Chapter Eight summarizes the work done in this research and draws conclusions. It also gives recommendations for further developments.

Some of the material presented in Chapters six and seven has been previously published in papers. In those cases where the candidate has been the author or a co-author of these papers, quotations are not indicated as such, but are simply referenced.

CHAPTER TWO

MEMS INERTIAL SENSORS - PERFORMANCE CHARACTERISTICS AND LIMITATIONS

The last two decades have witnessed an emerging trend of using MEMS in several navigation applications such as personal and land vehicular navigation. The low cost, lightweight, miniature size and modest power requirements of MEMS inertial sensors are the driving factors that attract many manufacturers especially in automobile applications. Improving the performance of MEMS-based inertial sensors is one of the main objectives of this research. This Chapter provides detailed background and analysis of MEMS sensors including different classification approaches as well as specific performance characteristics and limitations. In addition, since MEMS sensors are known for their high noise output, the commonly used methods for analyzing their noise characteristics are presented.

2.1 General Characteristics of Inertial Navigation Sensors

The operation of an inertial navigation system follows the laws of classical mechanics as formulated by Newton. Newton's first law states "*Every body continues in its state of rest, or uniform motion in a straight line, unless it is compelled to change that state by forces impressed on it*". Newton's second law states that "*Acceleration is proportional to the resultant force and is in the same direction as this force*" which translates to [Titterton and Weston, 1997]:

$$F = ma = mf + mg \quad (2.1)$$

where f is the acceleration produced by the force F other than the gravitational field. The acceleration a may be expressed as the total force per unit mass and g is the gravitational force. In case of a unit mass, Equation 2.1 becomes:

$$F = a = f + g \quad (2.2)$$

Given the ability to measure acceleration using an accelerometer, it is possible to calculate the change in velocity and position by performing successive mathematical integrations of the acceleration with respect to time. An INS usually contains three accelerometers, placed perpendicularly to one another, each of which is capable of detecting acceleration in a single direction.

In order to navigate with respect to the inertial reference frame, it is necessary to keep track of the direction(s) in which the accelerometers are pointing. Rotational motion of a body with respect to the inertial reference frame may be sensed using gyroscopic sensors and can therefore be used to determine the orientation of the accelerometers at all times. Given this information, it is possible to project the accelerations into the reference frame before the integration process takes place. Conventional gyroscopes make use of the inertial properties of a rotating wheel, or rotor, spinning at high speed [Savet, 1961]. Hence, inertial navigation is the process whereby the measurements provided by the accelerometers and gyroscopes are used to determine the position of the vehicle in which they are installed. By combining the two sets of measurements, it is possible to define the

translational motion of the vehicle within the inertial reference frame and hence to calculate the position within that frame [Cousins, 1990].

In order to ensure the appropriateness of an INS for a specific application, the application's performance requirements have to be fulfilled. The overall performance of an inertial navigation sensor is determined through accumulating the performances of individual inertial sensors involved in the system. This can be achieved in turn by quantifying the parameters or characteristics describing the performance of each particular inertial sensor through a series of lab tests. The most important among those characteristics are: bias, scale factor, output stability, thermal sensitivity, magnetic sensitivity, centrifuge, shock survivability, effect of vibrations, as well as ageing and storage [Titterton and Weston, 1997]. A brief description of each of these performance characteristics is given hereafter.

Bias

A sensor bias is always defined by two components: A deterministic component called bias offset which refers to the offset of the measurement provided by the sensor from the true input; and a stochastic component called bias drift which refers to the rate at which the error in an inertial sensor accumulates with time. The bias offset is deterministic and can be quantified by calibration while the bias drift is random in nature and should be treated as a stochastic process [El-Sheimy, 2002].

Scale factor

The scale factor is the relationship between the output signal and the true physical quantity being measured and it is usually expressed in parts per million (ppm). The scale factor is deterministic in nature and can be quantified or determined through lab calibration. The variation of the scale factor with the variation of the exerted acceleration/angular rate or temperature represents the scale factor stability and is usually called the non-linear part of the scale factor error [El-Sheimy, 2002].

Output Stability

The output stability of a sensor defines the run-to-run or switch-on-to-switch-on variation of the gyro-drift/accelerometer-bias as well as in-run variation of gyro-drift/accelerometer-bias. The run-to-run stability can be evaluated from the scatter in the mean output for each run for a number of runs given that the sensor is turned off then on again between each two successive runs. The in-run stability of a sensor is deduced from the average scatter of the measured drift in the output about the mean value during a single run [Andrews, 1975].

Thermal Sensitivity

Thermal sensitivity refers to the range of variation of the sensor performance characteristics, particularly bias and scale factor errors, with a change in temperature. A bias or scale factor correlation with temperature variation can be defined graphically or numerically (using a mathematical expression) through intensive lab thermal testing.

Such correlations can be stored on a computer for online use to provide compensation for temperature variation, provided a thermal sensor is supplied with the sensor.

Magnetic Sensitivity

Magnetic sensitivity refers to the influence that any external magnetic field might have on the drift characteristics of the sensor. Lab tests are performed to establish the magnitude of the dependence of the bias on the strength and orientation of an applied magnetic field [Smithson, 1987].

Centrifuge

A centrifuge provides a means for applying large steady or fluctuating accelerations to the sensor (accelerometer or gyroscope). The purpose of centrifuge evaluation is to investigate the response of the sensor to large acceleration and to establish its ability to withstand large continuous and fluctuating accelerations whilst the sensor is either operational or in a quiescent condition [Titterton and Weston, 1997].

Shock Survivability

The purpose of shock tests is to measure the response of a sensor to an applied shock and to establish the resilience of the sensor to such an applied acceleration over a very short duration, typically in the order of few milliseconds [Harris and Crede, 1961]. As in the case of a centrifuge, the sensor may be operational or motionless during the test. Comparisons with the mean value of the sensor bias before and after the application of the shock will indicate any transient or permanent change in the characteristics of the

sensor. In addition, the salient characteristics of the sensor before and after the application of the shock acceleration are compared.

Vibration Effect

Vibration tests normally represent the last series of tests carried out on the sensor owing to a potentially high risk of permanent damage to the sensor. Defining the vibration effect implies the following [Harris and Crede, 1961]:

- Determining the frequency at which any resonant responses of the sensor occur and their magnitude;
- Determining the anisoelasticity or acceleration squared (g^2) bias dependency of the sensor;
- Examining of the resilience and survivability of the sensor in a particular vibratory environment; and
- Estimating the change in noise characteristics of the output signal of the sensor experiencing a vibratory environment.

Aging and Storage

Many project applications require a sensor to have a shelf life of several years and still provide performance within its specifications at the end of that period. A simple method to evaluate ageing is to characterize a large number of sensors and then store them in a typical environment. Periodic withdrawal of a few, at say one or two years intervals, for re-evaluation should provide a guide to the form of ageing or change in performance of the sensor, thus allowing a comparison with the predicted changes [Kirkpatrick, 1974].

2.2 MEMS Sensors – Types and Operation Principles

There are different approaches to classifying MEMS sensors. They can be classified according to their method of detecting position of the mass. They may also be classified according to their mode of operation. Another approach is to classify them according to fabrication processes [Schwarz and El-Sheimy, 1999]. Each of these three classification approaches will be discussed next.

According to Mass Position Detection

Piezoresistive Sensors: This type of sensor senses the change of motion through the caused strain or deflection, which is proportional to the change in resistance of the Piezoresistive material. They are relatively inexpensive since they are bulk manufactured and use relatively simple interface electronics. They have also a wide frequency range. However, they have some disadvantages such as low precision and high sensitivity to temperature, which causes drifts in their output. In addition, using this type in a closed loop mode may increase complexity so they are always used in open loop for low cost and low precision applications [Hulsing, 1998].

Capacitive Sensors: The sensing element is a proof mass acting as the common centre electrode moving freely between two electrodes each forming a capacitor with the central mass. The deflection of mass from the centre caused by motion results in a differential change in capacitance of both electrodes. The output level of this type of sensor is about one order of magnitude better than the *Piezoresistive* type. Moreover,

they have excellent linearity and low sensitivity to temperature and are well suited to force rebalance designs usually needed in closed loop operations. On the other hand, they have some disadvantages like their great complexity, sensitivity to electromagnetic interference and small dynamic and frequency range [Varadan et al., 2001].

Piezoelectric Sensors: This type of sensor produces an electric charge when subjected to a force caused by the change of motion, or in other words, it transforms the physical work into an electrical signal. They have a relatively wide dynamic range. A disadvantage of those sensors is that they have low sensitivity to steady state inputs or small changes in motion, i.e., they only operate above a certain threshold [Smith and Balleto, 1994].

Resonant Element Sensors: This type of sensor measures the influence of the force caused by the change of motion on an under-damped mechanical structure vibrating at its resonant frequency. Such a resonant frequency changes in proportion with the acceleration.

There are also other types of sensors using various techniques for signal pick-off like tunnelling [Yeh and Najafi, 1997], cantilever beam [Smith and Balleto, 1994], heat source (or thermal), etc.

According to Operational Mode

Open Loop: The frequency difference in a pendulous system is sensed as a measure of the input acceleration. This type of sensor is simple to design and is cost efficient.

Closed Loop (with feed back): In this type of sensor, the displacement of the proof mass due to an applied acceleration is sensed and a restoring force is generated to counteract the motion [Zhou, 1998]. This type leads to increased complexity.

According to Fabrication

Polysilicon (surface) Micromachining: The sensing element is a thin silicon structure located on the surface of the die. The axis of sensitivity can be either perpendicular or parallel to the surface. Because of the multi-axis sensitivity, it is highly stable and a change in alignment is not expected [Hulsing, 1998]. It can be considered as an elegant solution that could be integrated with other electronic components on the same chip without changing the chip's fabrication process. This type of sensor fabrication is much closer to standard integrated circuit manufacturing than any of the previously discussed methods. The relatively low accuracy and small bandwidth represent important limitations for this fabrication process.

Bulk Micromachining: The sensing element is a single crystal inside a block that is sandwiched between two layers. The mass is assumed to move perpendicularly to the surface. This technique is mature and allows fabrication with well-defined parameters. The long-term stability of the device is enhanced due to the low stress of

the lightly doped layers and hence the entire mass is well protected. Bulk machined sensors are larger than surface machined, are relatively more expensive, and in addition are not as well suited for integration [Hulsing, 1998]. However, they overcome many of the accuracy limitations of polysilicon micromachining.

Table 2.1 summarizes the motion sensing technique as well as the advantages and the disadvantages of each category. Meanwhile, Table 2.2 shows the two main methods of MEMS-based inertial sensors fabrication processes [Abdel-Hamid et al., 2002].

Table 2.1 Classification of MEMS sensors according to mass position detection

Category	Motion sensing Technique	Advantages	Disadvantages
Piezoresistive signal pick-off	Deflection of the Piezoresistive material	Inexpensive, simple interface and wide range	Low precision and high sensitivity to temperature
Capacitive Sensors	The change in capacitance of two electrodes	More accurate, excellent linearity & lower temperature sensitivity	Great complexity, EM interference & small dynamic range
Piezoelectric Sensors	Produces an electric charge due to change in motion	Relatively wide dynamic range	Lower sensitivity to small changes in motion

Table 2.2 Classification of MEMS sensors according to fabrication processes

Category	Sensing Element	Advantages	Disadvantages
Polysilicon Micromachining	A thin silicon structure located on the surface of a die	Suites well for integration with other sensors	Relatively low accuracy and small bandwidth
Bulk Micromachining:	A single crystal inside a block sandwiched between two layers	Better accuracy than polysilicon micromachining	Larger size, more expensive & not good for integration

Beside the aforementioned technologies, which are commonly used in fabricating MEMS sensors, some other technologies have recently emerged. The latter technologies were known and used before in fabricating non-MEMS sensors (mechanical, solid state, vibratory inertial sensors, etc.). Nevertheless, thanks to the rapid advance in micromachining, these technologies have been successfully translated into the field of MEMS inertial sensors fabrication [Song, 1997]. The following are some of the recently produced types of accelerometers and gyroscopes based on these newly added technologies in micromachining:

Accelerometer Types

High-g Accelerometers: These accelerometers are designed to survive very high accelerations required in some applications. There is a niche market for accelerometers that read out very high g's for crash tests, impacts, and missile and artillery shell launches. For instance, the Charles Stark Draper Laboratory has

produced an accelerometer reading out 100,000 g's. Similar devices will almost certainly be commercialized in the future [Song, 1997].

Angular Accelerometers: MEMS angular accelerometers are used primarily to compensate for angular shock and vibration in disk read/write head assemblies. These devices, while similar to linear accelerometers in terms of design, fabrication, and readout, are designed with zero pendulosity (i.e., the centre of gravity is located at the centroid of the support springs) and are compliant to rotational motion yet stiff with respect to linear motion. Delphi and ST Microelectronics use capacitive MEMS sensors to manufacture angular accelerometers [Lutz et al., 1997].

Geophones: Geophones can be thought of as accelerometers with very high sensitivity and no DC output requirement. With no drift or bias stability specifications, geophone designs can be optimized to give the lowest noise floor. Applications for geophones include seismic sensing; machinery vibration and failure prediction, tracking and identification of vehicles or personnel, and underwater pressure gradient sensing. Conventional geophones incorporate permanent magnets and fine wire coils to measure velocity above their fundamental resonance [Song, 1997]. This is in contrast to capacitive accelerometers, which measure acceleration below their fundamental resonance.

Gyroscope Types

All MEMS gyroscopes take advantage of the Coriolis effect. In a reference frame rotating at angular velocity (Ω), a mass (m) moving with velocity (v) is subjected to a force [Yazdi et al., 1998] given by:

$$F = 2m v \times \Omega \quad (2.3)$$

Many types of MEMS gyroscopes are described in the literature, with most falling into the categories of tuning-fork gyros, oscillating wheels, Foucault pendulums, and wine glass resonators. Conventional (non-MEMS) spinning wheel gyros are common, but levitation and rotation of a MEMS device with no springs is yet to be commercialized [Yazdi et al., 1998].

Tuning Fork Gyroscopes: Tuning fork gyros contain a pair of masses that are driven to oscillate with equal amplitude but in opposite directions. When rotated, the Coriolis force creates an orthogonal vibration that can be sensed by a variety of mechanisms. An example for this type of gyro is the Draper Lab gyro [Bernstein et al., 1993], which uses comb-type structures to drive the tuning fork into resonance. Rotation causes the proof masses to vibrate out of plane and this motion is sensed capacitively. The resonant modes of a MEMS inertial sensor are extremely important. In a gyro, there is typically a vibration mode that is driven and a second mode for output sensing. In some cases, the input and output modes are degenerate or nearly so. If the I/O modes are chosen such that they are separated by $\sim 10\%$, the open-loop sensitivity

will be increased due to the resonance effect. It is also critical that no other resonant modes be close to the I/O resonant frequencies [Bernstein and Weinberg, 1996].

Vibrating-Wheel Gyroscopes: Many reports of vibrating-wheel gyros have been published [Putty, 1995]. In this type of gyro, the wheel is driven to vibrate about its axis of symmetry, and rotation about either in-plane axis. This results in the wheel's tilting, a change that can be detected with capacitive electrodes under the wheel.

Wine Glass Resonator Gyroscopes: A third type of gyro is the wine glass resonator. Fabricated from fused silica, this device is also known as a hemispherical resonant gyro. Researchers at the University of Michigan have fabricated resonant-ring gyros in planar form [Tang et al., 1997]. In a wine glass gyro, the resonant ring is driven to resonance and the positions of the nodal points indicate the rotation angle. The input and output modes are nominally degenerate, but due to imperfect machining some tuning is required. Silicon Sensing Systems, a joint venture between Sumitomo and British Aerospace, has brought to market an electro-magnetically driven and sensed MEMS gyro. A permanent magnet sits above the MEMS device. Current passing through the conducting legs creates a force that resonates the ring. This Coriolis-induced ring motion is detected by induced voltages as the legs cut the magnetic field.

Foucault Pendulum Gyroscopes: These devices are based on a vibrating rod that is typically oriented out of the plane of the chip [Geen, 1997]. They are therefore challenging to build with planar fabrication tools, but recent advances in MEMS technology allow very high aspect ratio MEMS that make it possible to fabricate the pendulum without hand assembly of the rod [Tang et al., 1997].

2.3 Specific Performance Characteristics of MEMS Sensors

A MEMS-based inertial sensor, as a member of the family of inertial sensors, can be also evaluated through verifying all the general performance characteristics described in Section 2.1. Nevertheless, some specific characteristics should be emphasized due to the special operational principles and fabrication processes of MEMS sensors. These specific performance characteristics include bandwidth, noise floor, cross-axis sensitivity, output stability (drift), linearity, dynamic range, shock survivability, and power consumption. Resonant frequency is also important because the sensor's upper useful frequency range is usually at a fraction of its resonant frequency, which also determines its sensitivity and displacement per (g) of acceleration. Nevertheless, the most important factor to consider is thermal sensitivity since it affects most of the other performance characteristics. In other words, with the change of the sensor's temperature, the parameters describing most of the other sensor characteristics will definitely vary significantly. In the following three subsections, the most significant three characteristics, describing the MEMS inertial sensors performance, are emphasized. These three characteristics are the output stability, thermal sensitivity and noise characteristics.

2.3.1 Output Stability

The stability of inertial sensors output refers to the variation of their output errors (mainly bias and scale factor) either during the same run, namely in-run stability, or from turn on to another and, in this case, it is called run-to-run stability. The stability characteristics should be studied very carefully in the case of MEMS inertial sensors since, unlike other tactical or navigation grade sensors, they are known for their high output drift with time as well as their poor run-to-run stability [Salychev, 1998]. This is mainly because MEMS sensors, due to their miniature size and fabrication process, are highly sensitive to changes in the surrounding environmental conditions.

The in-run instability of MEMS sensor biases is normally treated as a random process and can be expressed using different algorithms. Such algorithms include the Gauss-Markov algorithm with its different orders [Gelb, 1989], random walk, autoregressive (AR) model [Pimbley, 1992; Nassar, 2003], etc. The run-to-run instability error of MEMS sensors is considered as a random constant during a single run and can be expressed by the scatter of the mean output for each single run of a number of runs [Titterton and Weston, 1997].

2.3.2 Thermal Sensitivity

The term “thermal sensitivity” is defined as the change in the different parameters defining the performance of the inertial sensor due to the variation in the sensor’s ambient temperature [Titterton and Weston, 1997]. It can be inferred from the latter definition that thermal variation is a dominant factor that impacts most of the other

performance characteristics of the sensor. Due to the miniature design and lightweight of MEMS sensors, the expected effect of temperature change on the sensor characteristics, and particularly bias components, should be significantly higher if compared to other non-MEMS sensors. This necessitates the compensation for thermal variation of performance parameters in order to avoid a significant drift in the sensor output, which will definitely affect the overall accuracy of navigation systems using MEMS-based inertial sensors.

The thermal variation of MEMS inertial sensors affects both the deterministic characteristics such as biases and the stochastic characteristics such as noise outputs. With respect to the deterministic effect, the expansion or shrinkage of the sensor frame will result in a different height (H) of the two pairs of pivot points with respect to the frame of the sensor. Even using equal length pendulum arms would produce a tilt of the test or proof mass and hence, a drift Δx of its equilibrium position is described as [Bertolini, 2000]:

$$\Delta x = \alpha \frac{H}{S} \Delta T \quad (2.4)$$

where α is the thermal expansion coefficient, ΔT is the change in temperature and S is the distance between the pendulum arms. Chapter Six provides a detailed study of the thermal variation of MEMS accelerometer biases as well as MEMS gyro drifts. In addition, a convenient method to model this (non-linear) thermal variation appropriately is also proposed and verified.

On the other hand, temperature contributes to the sensor noise with a term called thermal noise. This contribution is given later in the following subsection (Equations 2.6 and 2.7) as Brownian motion of the proof mass as well as Brownian equivalent acceleration.

2.3.3 Stochastic (Noise) Characteristics

MEMS-based inertial sensors are known to exhibit relatively high output noise accompanying output signals or measurements. This noise results in short-term errors that may coincide in the frequency ranges with those of the true motion of interest. Therefore, the separation of these short-term errors from the motion signal is difficult and risky when using traditional filtering like low-pass and high-pass filters.

There are many contributors to noise in the sensor itself such as the readout electronics, mechanical damping, and electrical resistances. MEMS sensors are so small that the noise of the devices' mechanical resistance must be considered whereas it is usually ignored in larger sensors. Just as Brownian motion excites bacteria and dust motes, it can be a large force on a tiny MEMS component. The Brownian force [Gabrielson, 1993] is given by:

$$F_B = \sqrt{4KTD} \quad (2.5)$$

which causes Brownian motion of the proof mass, X_B :

$$X_B = \sqrt{\frac{4KTD}{K_{sp} + j\omega D - \omega^2 M}} \quad (2.6)$$

where:

D is the damping coefficient of proof mass M supported by spring constant K_{sp} , K is Boltzmann's constant and T is the ambient temperature.

Solving for the acceleration that generates the same motion, X_B , and substituting:

$$Q = \omega_o M / D, \quad \omega_o = \sqrt{K_{sp} / M}, \quad \text{and} \quad g = 9.8 \text{ m / s}^2$$

gives the Brownian equivalent acceleration noise in g/ Hz:

$$g_{n,B} = \frac{\sqrt{4KTD}}{M g} = \frac{1}{g} \sqrt{\frac{4KT\omega_o}{M Q}} \quad (2.7)$$

From Equation 2.7, we see that a large mass M and a high Q (low damping) help achieve a low noise floor. To achieve a large mass in a micro-machined sensor typically requires a wafer-thick proof mass carved out of the sensor chip. For absolute minimum noise, the damping constant must be reduced by suspending the proof mass in a vacuum from purely elastic springs. Feedback prevents the sensor from ringing at its resonant frequency [Bernstein et al., 1998].

Flicker noise is another noise source in MEMS devices due to current flow. This noise phenomenon is a direct function of the channel dynamics experienced by bodies carrying the sensor. Such noise is expressed as [Pottenger and Kaiser, 2001]:

$$v_{Flicker}^2 = \frac{K_f}{C WL} \frac{1}{f} \quad (2.8)$$

where K_f is a process-dependant parameter indicating the cleanliness of the fabrication process, WL represents the sensor geometry, C is a constant and f is the frequency of the affecting dynamics.

Flicker and thermal noise components create noise spectra with flicker noise dominating at the low frequencies. At greater frequencies, the thermal noise is the dominant mechanism [Pottenger and Kaiser, 2001].

2.4 Methods for Analyzing the Noise Characteristic

Random (noise) errors characterization of MEMS inertial sensors provides the necessary noise information for the integrated sensors assembly (ISA) and the subsequent inertial measuring units based on these sensors. Generally, noise imposed on the inertial sensor outputs can be modeled as a stochastic process using time domain or frequency domain analysis techniques for its characterization. These techniques involve the computation of the power spectral density function (PSD), the correlation function, and the Allan variance [Aly et al., 2004].

Noise contributors in typical gyros/accelerometers outputs include white angular (velocity) noise, quantization noise, white rate (acceleration) noise, correlated random noise, bias instability (1/f or flicker noise) and sinusoidal noise [IEEE, 1998]. The power spectral density function (PSD) and the Allan variance methods have more general applications than the investigation of stochastic models. They have been adopted as preferred means of analysis by the inertial systems community. One of the most notable

deficiencies of the standard Allan variance is the non-unique characterization of white, flicker, angle (velocity) and quantization noise at the high frequency end and random rate ramp at the low frequency end [IEEE, 1988]. However, these may be sorted out by using the modified Allan variance, pre-whitening, or pre-filtering the reading data. Pre-filtering is essential to remove external noise sources or added noise during the acquisition process.

2.4.1 Frequency Analysis of Inertial Signal

PSD is the most commonly used representation of the spectral decomposition of a time series. It is a powerful tool for analyzing and characterizing data, and for stochastic modelling. The Allan variance time domain technique has been originally developed for the characterization, estimation, and prediction of precision clocks and oscillators in the time domain. Due to its nature, it had been employed to characterize a inertial sensors [Tehrani, 1983; IEEE, 1998]. The different types of noise associated with inertial sensors can be revealed via the PSD and/or Allan variance method. As these types of noise errors possess different frequency dependence, a typical PSD or Allan variance plot would contain these types of noise revealed at different frequencies or cluster times (t). It should be noted that the uncertainty in the data is assumed to be generated by a noise source of specific character. Source covariance is then estimated from the data. It should be noted that the Allan variance technique can be used as a complementary method to any of the other frequency domain analysis techniques [IEEE, 1997]. See Hou and El-Sheimy [2003] for more details about using the Allan variance and modified Allan variance in modelling inertial sensors errors.

2.4.1.1 Fourier Transform (FT) analysis

Consider N samples of the sensor output with sample rate Δt given by:

$$x_0, x_1, x_2, \dots, x_{N-1}. \quad (2.9)$$

Thus, the length of the time ensemble is $T = N \Delta t$. The autocorrelation function of a stationary stochastic process $x(\tau)$ is defined as:

$$\phi(\tau) = E\{x(t + \tau) \cdot x(t)\}, \quad (2.10)$$

where E denotes the expectation of the process and τ is the correlation time. It had been shown that [IEEE, 1998] choosing $\tau \ll T$ results in an un-biased estimate of the correlation function. An estimate to the two-sided *PSD* can be obtained by taking the Fourier transform of the estimate of the auto-correlation function:

$$\Phi(\omega) = FT(\phi(\tau)), \quad (2.11)$$

where *FT* denotes the Fourier transform and ω denotes the angular frequency. In the following computations, however, the one-sided *PSD* estimate will be obtained from the following equation:

$$\Phi^1(\omega) = \frac{1}{T} |X(\omega)|^2, \quad (2.12)$$

where $X(\omega)$ is the Fourier transform of the measured time series $x(t)$ and the superscript is used to differentiate between the normal (two-sided) PSD and the single sided PSD. The discrete *FT* approximates the continuous *FT* at discrete frequencies $\omega_j = 2\pi f_j$ where f_j is given by:

$$f_j = \frac{j}{N \Delta t} = \frac{j}{T} [\text{Hz}] \quad (2.13)$$

2.4.1.2 Different Noise Components and their Signature

Normally, the *PSD* of a random process is expected to exhibit even order log – log slopes of $-2, 0, +2, \dots$, indicating even powers of frequency. Thus, the different types of noise imposed on the sensor reading are represented in the *PSD* by straight lines with different slopes. These noise types can be summarized as follows [IEEE, 1998]:

i) Rate (acceleration) random walk:

This is a random process of uncertain origin, possibly a limiting case of an exponentially correlated noise with a very long correlation time. The rate (acceleration) *PSD* associated with the noise is given by:

$$\Phi^1(\omega) = 2 \cdot \frac{k^2}{\omega^2}, \quad (2.14)$$

where k is the rate random walk coefficient. The log-log plot of the *PSD* for rate random walk has a slope of -2 .

ii) Angle (Velocity) Random Walk:

The main source of this error is the spontaneous electronic vibrations inside the sensor. Other high frequency noise terms that have a correlation time much shorter than the sample time can also contribute to the gyro angle or accelerometer velocity random walk.

These noise terms are all characterized by white spectrum. The associated rate noise *PSD* is given by:

$$\Phi^1(\omega) = 2 \cdot N^2, \quad (2.15)$$

where N is the angle (velocity) random walk coefficient. Typical log-log plot of the PSD for this type of noise has a zero slope.

iii) Bias Instability

The origin of this noise source is the electronic or other components susceptible to random flickering. Because of its low frequency nature, it shows up as bias fluctuations in the data. The rate *PSD* associated with this noise is:

$$\Phi^1(\omega) = \begin{cases} 2(B^2 / \omega), & \omega \leq \omega_o \\ 0, & \omega > \omega_o \end{cases}, \quad (2.16)$$

where B is the bias instability coefficient and ω_o is the cut off frequency. Bias instability shows up in log-log plot of the PSD as a slanted line with a slope of -1 for frequencies below the cut-off frequency and zero slope at higher frequencies.

iv) Exponentially correlated (Markov) noise:

This noise is characterized by an exponentially decaying function with a finite correlation time. The rate PSD for such a process is:

$$\Phi^1(\omega) = 2 \cdot \frac{(q_c T_c)^2}{1 + (\omega T_c)^2}, \quad (2.17)$$

where q_c is the noise amplitude, and T_c is the correlation time. For $\omega < T_c^{-1}$ the PSD is represented by a zero-slope line, while for $\omega > T_c^{-1}$, the PSD of the Markov correlated noise has a slope of -2 . The PSD behaviour of all mentioned types of noise is shown in Figure 2.1. Figure 2.2 and Figure 2.3 shows the PSD of real outputs (in static mode) of two MEMS-based Analog Device gyroscopes, namely the CRS03 and CRS02 models, respectively. Inspection of the latter two plots reveals the presence of a rate random walk, bias instability, angle random walk, and correlated Markov-type noise.

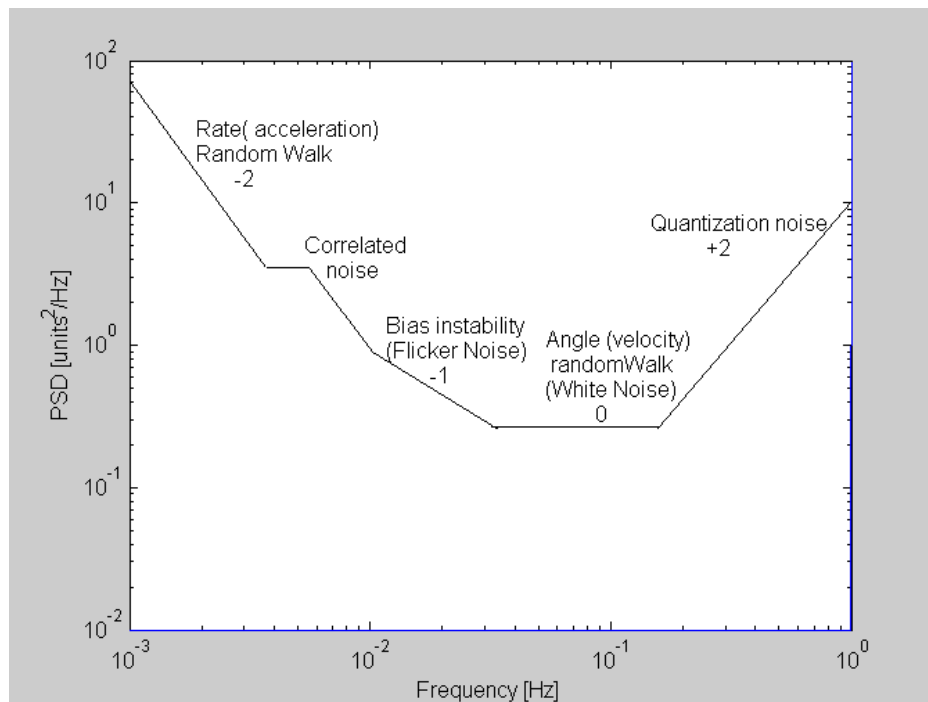


Figure 2.1 Typical slopes in $\log_{10} - \log_{10}$ of one-sided PSD versus frequency (accelerometer acceleration data or gyroscope angle rate data).

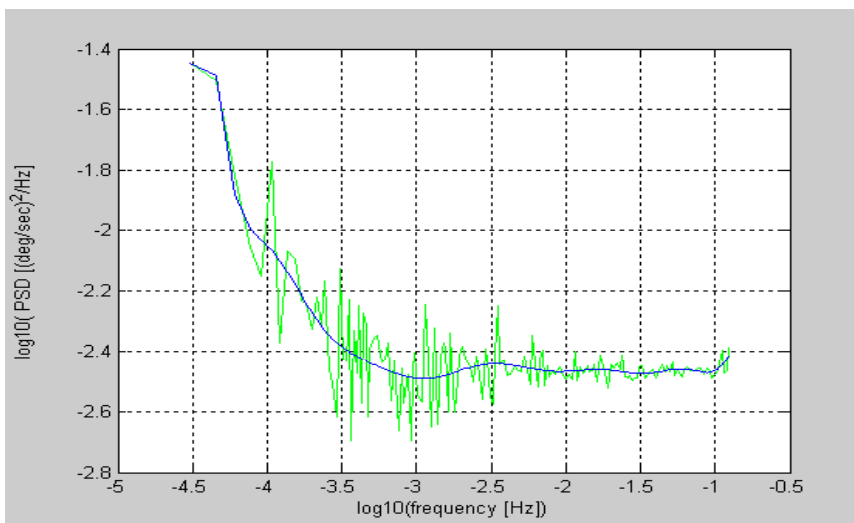


Figure 2.2 Ensemble and frequency average of the one-sided PSD for the MEMS gyro CRS03

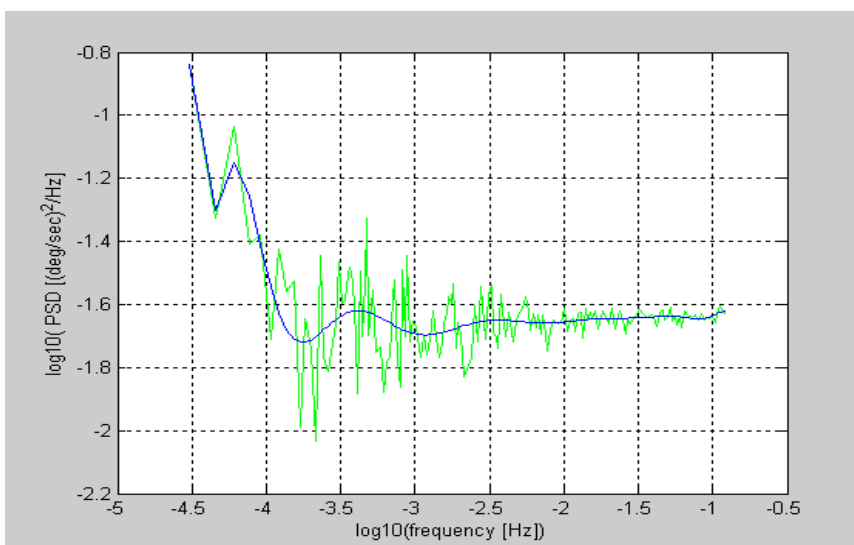


Figure 2.3 Ensemble and frequency average of the one-sided PSD for the MEMS gyro CRS02

2.4.2 Time Analysis of Inertial Signals

The noise components of MEMS sensors outputs can be treated as a random process. This in turn may be thought of as a collection or ensemble of functions of time such that any one of them might be observed on any trial of an experiment [Gelb 1989]. There are common random process classes that are usually used to describe the behaviour of noise components in most inertial applications. Such classes start from white noise, the simplest to the different orders of Gauss-Markov processes, which mainly describes the correlated (colored) noise. Each of the previous classes has its own shape or sign of power spectral density and autocorrelation function through which it can be identified or recognized.

2.4.2.1 Auto-Correlation Function and Correlation Time

Correlation functions may be introduced as joint expectations for different processes x and y and termed cross-correlation function [Merhav, 1996]:

$$E(x_{t_1}, y_{t_2}) = \frac{1}{n_s} \sum_{k=1}^{n_s} x_k(t_1) y_k(t_2) \quad (2.18)$$

If $x \equiv y$, we refer to it as the autocorrelation function. Then the definition of the autocorrelation function $C_{xx}(t_1, t_2)$ can be expressed as:

$$C_{xx}(t_1, t_2) \triangleq E(x_{t_1} x_{t_2}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \alpha_1 \alpha_2 p_{x_{t_1} x_{t_2}}(\alpha_1, \alpha_2) d\alpha_1 d\alpha_2 \quad (2.19)$$

The function is the expectation of x_{t_1} , at t_1 , being in the slice α_1 , $\alpha_1 + d\alpha_1$ and x_{t_2} , at t_2 , being in the slice α_2 , $\alpha_2 + d\alpha_2$. This statistical descriptor is a measure of the statistical

similarity of $\{x\}$ with itself at the instants t_1 and t_2 given by $\tau = t_1 - t_2$. If the process $\{x\}$ varies slowly, this similarity or correlation will be high. Conversely, if it varies rapidly the correlation will be low. The correlation time (τ) is the time when the signal is expected to repeat itself. Determining the value of correlation time is vital to model the stochastic characteristics (correlated noise) of a variable, as well be explained in the following sub-section [Merhav, 1996]. Figure 2.4 represents a typical autocorrelation function of a Gauss-Markov random process (see Section 2.4.2.2). In addition, the autocorrelation function of the real output sequence of the Cross-Bow AHRS300CA Z-axis MEMS gyroscope, under different temperatures, is shown in Figure 2.5. The upper row of the latter figure shows the autocorrelation functions of the raw output while the lower row illustrates the same functions but after decomposing the raw output using the WMRA. For more details on WMRA, see Chapter Four. The figure clearly indicates that the autocorrelation functions became much smoother after applying WMRA de-noising. This makes them more appropriate to decide on the proper noise model (e.g. the proper order of the Gauss-Markov process).

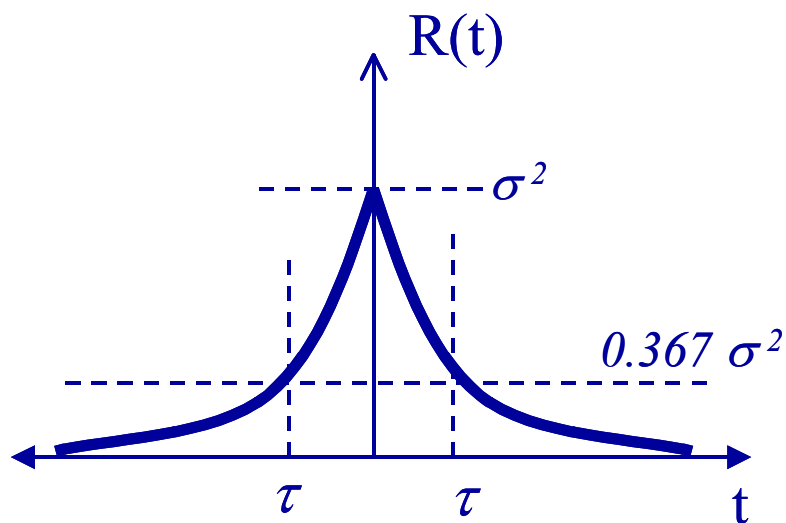


Figure 2.4 Autocorrelation function of a 1st order Gauss-Markov random process and its correlation time

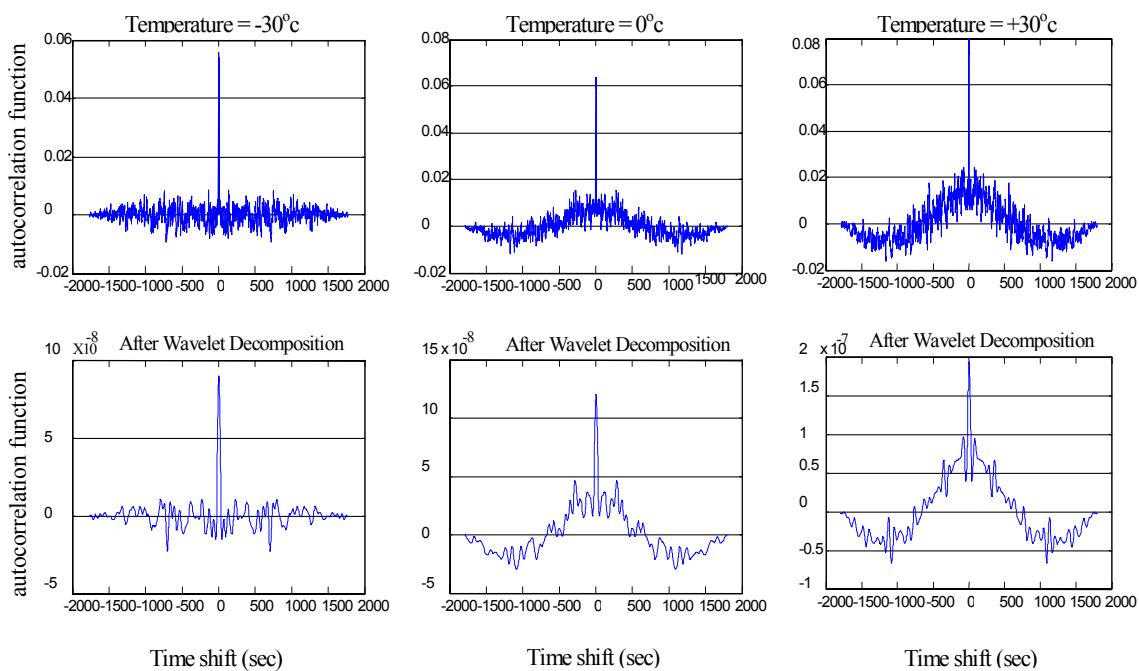


Figure 2.5 Autocorrelation functions of the Cross-Bow AHRS300CA Z-gyro at three different sensor temperatures [-30°, 0° and +30°C]

2.4.2.2 Modeling Correlated Noise

Based on the autocorrelation sequence of the random process, noise can be classified into different types as well be illustrated hereafter.

White Noise has a particularly simple (constant) power spectral density, which means that the power is uniformly distributed over all frequency components (ω) in the full infinite range [Gelb, 1989]. The power spectral density Φ_{xx} and the autocorrelation function φ_{xx} of white noise can be expressed as (see Figure 2.6):

$$\Phi_{xx}(\omega) = \Phi_o = \text{const.} \quad (2.20)$$

$$\varphi_{xx}(\tau) = \Phi_o \delta(\tau) \quad (2.21)$$

where τ is the time shift and $\delta(\tau)$ is the Dirac function, which has a value of infinity at the zero time shift and zero at any other value of (τ). Consequently, this means that the value of the autocorrelation function of white noise is just a pulse at ($\tau = 0$)

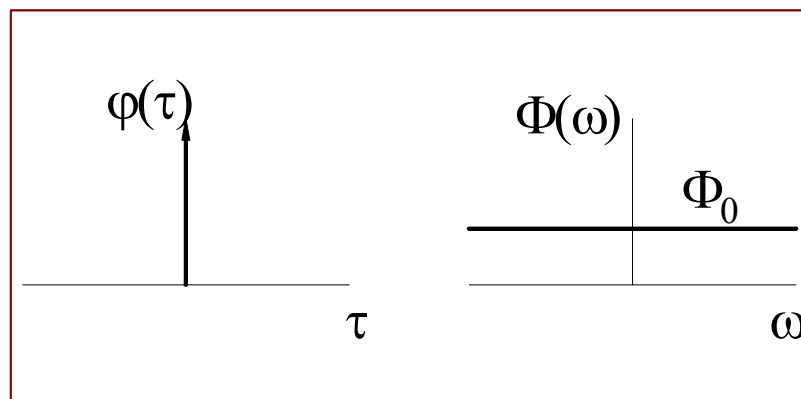


Figure 2.6 Power spectral density and autocorrelation function of white noise

Gauss-Markov Processes constitute a special class of random processes that can be generated by passing white noise through simple shaping filters. As a result we get an infinite number of random process orders called the Gauss-Markov processes family. The power spectral density and the autocorrelation function of the first order Gauss-Markov have their own shapes (see Figure 2.7) and can be expressed as [Gelb, 1989]:

$$\Phi_{xx}(\omega) = \frac{2\beta_1\sigma^2}{\omega^2 + \beta_1^2} \quad (2.22)$$

$$\varphi_{xx}(\tau) = \sigma^2 e^{-\beta_1|\tau|} \quad (2.23)$$

where β_1 is the inverse of the correlation time and σ^2 is the variance of the random process. In this case if the time shift τ is taken to be equal to the correlation time, Equation 2.23 becomes:

$$\varphi_{xx}(\tau) = \sigma^2 e^0 = 0.367 \sigma^2 \quad (2.24)$$

This will give the value of the autocorrelation function corresponding to the correlation time as shown in Figure 2.4.

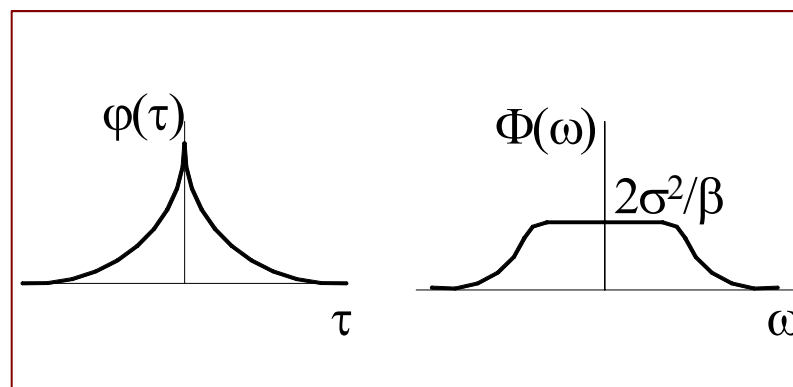


Figure 2.7 Power spectral density and autocorrelation function of first order Gauss-Markov Random Process

The second order Gauss-Markov process has a different shape power spectral density and autocorrelation functions (see Figure 2.8) that can be expressed as [Brown and Hwang, 1985]:

$$\Phi_{xx}(\omega) = \frac{4\beta_2^3\sigma^2}{(\omega^2 + \beta_2^2)^2} \quad (2.25)$$

$$\varphi_{xx}(\tau) = \sigma^2 e^{-\beta_2^2|\tau|} (1 + \beta_2|\tau|) \quad (2.26)$$

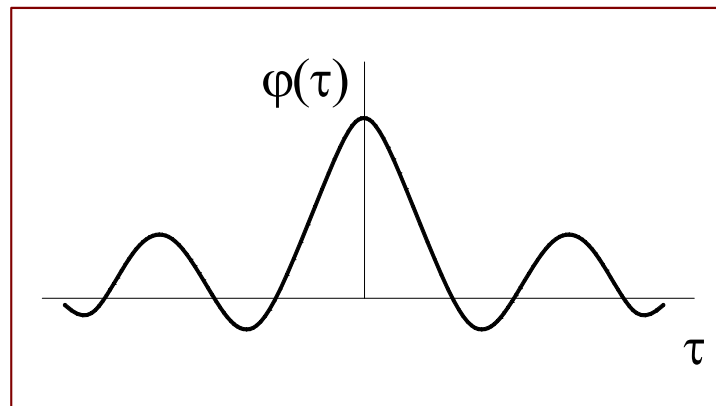


Figure 2.8 Autocorrelation function of a second order Gauss-Markov Random Process

Each of the higher order Markov processes has its own shape and functional model. When the order of the process reaches infinity the result is a bias and the corresponding power spectral density and autocorrelation function are (see Figure 2.9):

$$\Phi_{xx}(\omega) = 2\pi\sigma^2\delta(\omega) \quad (2.27)$$

$$\varphi_{xx}(\tau) = \sigma^2 \quad (2.28)$$

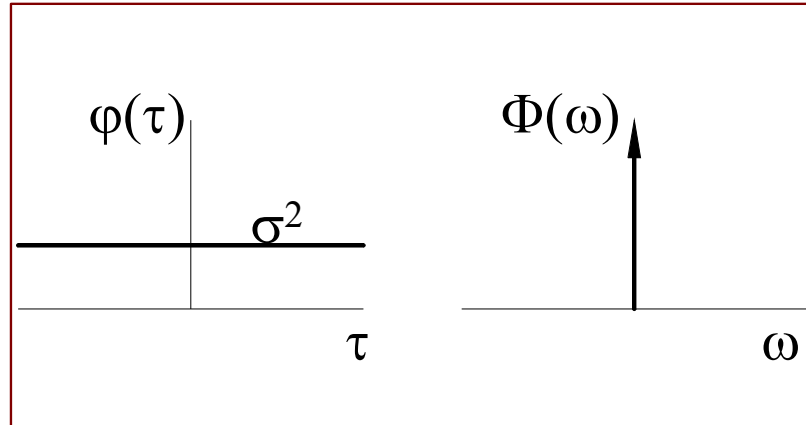


Figure 2.9 Power spectral density and autocorrelation function of infinite order Gauss-Markov Random Process (bias)

As a final remark, white noise can be thought of as a zero-order Gauss-Markov random process since there is no correlation between the probability distribution of the current epoch random process and any of the previous epochs' ones [Brown and Hwang, 1985].

Either of the different orders of Gauss-Markov random processes, discussed in this section, can be implemented to model the stochastic bias drift of an accelerometer or gyroscope. However, the first order Gauss-Markov is the most commonly used due to its simplicity. Stochastic modeling of the bias drift is a vital step to create a complete dynamic error model in KF solution of the integrated INS/GPS system. This will be discussed in details in the following chapter.

CHAPTER THREE

KALMAN FILTERING OF INS/GPS INTEGRATED SYSTEMS

In this Chapter, the integration of an INS and GPS will be discussed from the estimation viewpoint. The error characteristics of both systems and their complementary nature will be addressed first. Then, the algorithms used in the INS/GPS conventional (Kalman) integration approach will be described. Finally the advantages and shortcoming of this Kalman integration approach will be discussed.

3.1 GPS and INS Error Characteristics

Integrated navigation systems are gaining adoption in positioning and navigation applications because they allow for implementation of data fusion concepts at the measurement level. The integration of navigation technologies such as the GPS and INS provides superior performing systems in comparison with either a GPS or an INS stand-alone system. In addition, the complementary nature of the two systems fosters the development of systems that integrate both technologies [Mohamed, 1999]. In georeferencing applications, for instance, GPS-derived positions and velocities are excellent external measurements for updating the INS and providing the imaging sensors with position parameters, thus improving the applications' long-term accuracy. Similarly, the INS can provide precise position and velocity data for GPS signal acquisition and reacquisition after outages as well as the orientation parameters for the vision-based subsystems. The integration of GPS with INS is commonly implemented using KF in

different schemes such as loosely, tightly, and ultra-tightly coupled integration and can be implemented either in a closed-loop or open loop framework [Phillips and Schmidt, 1996].

3.1.1 GPS Observation Measurements Errors

In an INS/GPS integrated system, the GPS is responsible for providing the update measurements. It provides three different observations, namely, pseudo-code, carrier phase, and Doppler (or phase rate). Through the latter observation, velocity update measurements can be extracted. On the other hand, the position update measurements can be derived by solving either pseudo-code or carrier phase observations from at least four satellites. For precise kinematic applications, however, carrier phase measurements are generally used. The pseudorange measurement in single point positioning (SPP) mode is expressed as [Lachapelle, 2004 ; Hoffmann-Wellenhof et al., 1997]:

$$P = \rho + d\rho + c(dt - dT) + d_{ion} + d_{trop} + \xi_{Pmult} + \xi_{Pnoise} \quad (3.1)$$

where,

P	measured pseudorange
ρ	geometric spatial distance between the receiver and the satellite
$d\rho$	satellite orbital error
c	speed of light
dt	satellite clock offset
dT	receiver clock offset
d_{ion}	ionospheric signal delay

d_{trop}	tropospheric signal delay
$\xi_{P\ mult}$	pseudorange signal multipath
$\xi_{P\ noise}$	pseudorange signal noise.

The carrier phase observation equation, in SPP mode, can also be written as [Parkinson and Spilker, 1996; Lachapelle, 2004]:

$$\Phi = \rho + d\rho + c(dt - dT) + \lambda N - d_{ion} + d_{trop} + \xi_{\Phi\ mult} + \xi_{\Phi\ noise} \quad (3.2)$$

where,

Φ	observed carrier phase (measurement)
ρ	geometric spatial distance between the receiver and the satellite (observable)
$d\rho$	satellite orbital error
c	speed of light
dt	satellite clock offset
dT	receiver clock offset
λ	carrier wavelength
N	initial carrier ambiguity
d_{ion}	ionospheric signal delay
d_{trop}	tropospheric signal delay
$\xi_{\Phi\ mult}$	carrier phase signal multipath
$\xi_{\Phi\ noise}$	carrier phase signal noise.

Errors contributing to the phase observable in single point positioning mode (SPP) are due to satellite orbital perturbations and the satellite clock bias, the receiver clock bias and measurement noise, the signal delay through the ionosphere and the troposphere; and signal multipath. These error sources impact the positioning accuracy as the final product of phase observable processing. The largest error is due to the receiver clock. The next significant error source is the medium in which the signal must travel. This includes the ionosphere, which is the part of the atmosphere between about 50 km and 1000 km and has many free electrons; and the troposphere, which is a non-depressive medium and contains mostly electrically neutral particles. The anticipated overall positioning error in this case could reach several metres (10-20 metres) [Wells et al., 1986; Parkinson and Spilker, 1996; Lachapelle, 2004]. Because the errors incurred in SPP mode cannot meet the accuracy requirements, SPP mode is not suitable for precise applications.

To eliminate or reduce measurement errors, differential observables are used. In real-time kinematic applications, single differencing between receivers is often used to reduce measurement errors and eliminate the satellite clock error. This approach, however, does not eliminate the receiver clock error. Since the receiver clock drifts over time, a new receiver clock term has to be modeled for each observation epoch. The estimation of the receiver clock term requires precise modeling, which is not always achievable especially for quartz clocks commonly used in GPS receivers. In precise kinematic applications, double differencing is often used to eliminate receiver and satellite clock errors and reduce correlated measurement errors. Double differencing is carried out by differencing

between receivers and between satellites at the same observation epoch. The double differenced phase observation equation takes the form:

$$\Delta\nabla\Phi = \Delta\nabla\rho + \Delta\nabla d\rho + \lambda\Delta\nabla N - \Delta\nabla d_{ion} + \Delta\nabla d_{trop} + \xi_{\Delta\nabla\Phi mult} + \xi_{\Delta\nabla\Phi noise} \quad (3.3)$$

where $\Delta\nabla$ represents the double differencing operator.

In the double differenced observables, the receiver and satellite clock offsets are eliminated, the correlated measurement errors are reduced, but however, the phase signal noise is amplified. The phase noise is a function of the receiver tracking bandwidth and is usually within one percent (1%) of the phase signal wavelength [Lachapelle, 2004; Parkinson and Spilker, 1996]. For the GPS L1 signal, the phase noise corresponds to about 2 mm (1 σ), which will be doubled after double differencing. Due to the high geometric correlation of range measurements between the receivers and the satellites over short baselines, the orbital error is considerably reduced by differencing. For a baseline of twenty 20 km in length, the maximum expected range error due to $\delta\rho$ is 20 mm with a relatively slow variation.

Some of the ionospheric delay may be removed by using the broadcast Klobuchar model [Klobuchar 1991]. Based on the state of the ionosphere, the ionospheric delay can be determined and subsequently totally eliminated by combining L1 and L2 phase observations using dual frequency receivers. This combination, however, is not recommended for short baselines as it results in a three times noisier observables than the double differenced observables [Lachapelle et al., 1987 ; Langley, 1993]. It also destroys

the integer nature of the phase ambiguity of the resulting observables. For short baselines, the ionospheric delay is spatially and temporally correlated; the spatial correlation is found to extend from about 200 km up to 1000 km, and the temporal correlation ranges from several minutes to hours under disturbed and normal conditions, respectively [Wild et al, 1990]. This means that the major part of the ionospheric delay is removed by differencing. The remaining part of the ionospheric delay is normally less than ten percent (10%) of the signal wavelength [Lachapelle et al., 1992] or about one to two ppm of the baseline length [Lachapelle, 2004]. For instance, for a double differenced L1 phase observables over a 10 km baseline, the maximum expected residual ionospheric delay is 20 mm.

The effect of the neutral or non-ionized part of the atmosphere is denoted as the tropospheric delay. The neutral atmosphere is a non-depressive medium with respect to radio waves with frequencies up to 15 GHz [Bauersima, 1983]. Thus, it identically delays both L1 and L2 signals. Consequently, dual frequency receivers that are capable of removing ionospheric delays cannot remove the tropospheric effect like they do with the ionospheric one. The neutral atmosphere (troposphere) shows a maximum effect at the horizon and a minimum one at the zenith. Up to ninety percent (90%) of the tropospheric delay is due to the so-called dry component and the rest is due to the so-called wet component. The dry component can be predicted with high accuracy using empirical models [Wells et al., 1986 ; Lachapelle, 2004]. The wet component effect, on the other hand, is much more difficult to model because of the strong variations of water vapour with respect to time and space. However, for short baselines, the tropospheric residual

error almost disappears by using double differenced measurements and rarely exceeds ten 10 mm for land-based applications. However, for airborne applications where the height separation between the reference station and the rover can easily reach 1 km, a refined tropospheric model becomes important [Seeber 1993].

Multipath is mainly caused by reflecting surfaces near the receiver. Its impact on the measurements is totally uncorrelated between stations and therefore cannot be removed by differencing. However, at the same station, a multipath signature repeats itself every sidereal day because of the repeated satellite-receiver geometry. The maximum multipath effect of phase measurements does not exceed quarter wavelength, i.e. 50 mm for L1 observations [Braasch 1998]. Temporal correlation between 1 to 20 minutes has been reported depending at observation sites [El-Rabbany 1996]. Multipath is considered a limiting factor when ambiguities are to be resolved within a few epochs. Modern receivers usually employ multipath mitigation techniques to reject the reflected signal. In general, multipath has less effect on kinematic applications than on static applications because multipath is more random in the latter case.

Many of the aforementioned sources of GPS measurement errors, particularly receiver clock errors may degrade the SPP accuracy. In order to achieve improved accuracy and eliminate or at least reduce the effect of such errors, differential GPS (DGPS) particularly with the double differenced technique is preferably implemented. Double differencing is based on the use of two or more receivers where one (stationary) reference or base receiver is located at a known location while the position of the other (stationary or

moving) remote receiver is to be determined. The known position of the reference receiver is used to estimate corrections to the GPS derived position or to the observed ranges. These corrections are then transmitted to the remote receiver to compute the remote position with far more accuracy than in the SPP mode. The higher accuracy is based on the fact that GPS errors are very similar over a distance of up to several hundred km. In several navigation applications, users are primarily interested in real time accuracy at the metre level, which cannot be achieved by SPP but can be obtained by DGPS. With the application of the DGPS technique based on C/A code ranges, a few metre-level accuracy can be achieved. If a sub-metre level accuracy is required, phase smoothed code ranges or high performance C/A code receivers must be used [Lachapelle et al., 1992]. An even higher accuracy level can be reached when carrier phase measurements are used. For instance, a sub-decimetre accuracy level in real time can be obtained with baseline ranges of up to 20 km [DeLoach and Remondi, 1991].

Velocity information which is derived from GPS Doppler or phase rate observables is less prone to errors. This is due to the fact that double differenced Doppler observables display similar error characteristics as of phase observables and velocity information is implicitly obtained from differencing over short periods of time during which many errors are constant. In general, they are less sensitive to orbital and atmospheric effects. They are not subject to cycle slips and have no initial ambiguity. The noise level of the double differenced Doppler observables for modern geodetic GPS receivers is usually below 10 mm/s [Lachapelle et al., 1992].

3.1.2 INS Observation Measurements Errors

The INS in an integrated INS/GPS system is responsible for interpolating position between updates (short-term positioning component), as well as providing attitude information. The INS system comprises two types of inertial sensors, namely accelerometers, which measure incremental linear velocities and gyroscopes, which measure incremental angular velocities. At the accelerometer sensor level, a linear acceleration measurement can be described in a simplified form as follows [El-Sheimy, 2002; Schwarz and Wei, 2000]:

$$\ell_a = f + b + (S_1 + S_2)f + N_f + \gamma + \delta g + \xi_f \quad (3.4)$$

where,

ℓ_a is the acceleration measurement

f is the specific force (observable)

b is the accelerometer bias

S_1 is the diagonal matrix of linear scale factor errors

S_2 is the diagonal matrix of non-linear scale factor errors

N_f is a matrix representing accelerometer triad nonorthogonality

γ is the vector of normal gravity

δg is the anomalous gravity vector

ξ_f is the accelerometer measurement random noise.

In a similar approach, the angular velocity measurement can be expressed as:

$$\ell_{\omega} = \omega + d + S_{\omega} + N_{\omega} + \xi_{\omega} \quad (3.5)$$

where,

ℓ_{ω} is the angular velocity measurement

ω is the angular velocity observable

d is the gyroscope instrument bias

S_{ω} is the diagonal matrix of gyroscope scale factor errors

N_{ω} is a matrix representing gyroscope triad nonorthogonality

ξ_{ω} is the gyroscope measurement random noise.

The measurement bias is usually modeled as a constant bias in the accelerometer or the gyroscope readings. This means that the derived velocity errors and attitude errors are linear with time. Both biases are the result of the manufacturing imperfection of the sensors. The bias value is determined by calibrating the system in the lab or during field tests. Measurements are then compensated for the estimated bias. In reality, however, these biases are never completely constant. Their time-variable part is modeled as stochastic noise.

The scale factor is the ratio between the sensor input and output. It depends on the system dynamics and the temperature variation. Under normal operational conditions and relatively low dynamics, the scale factor remains constant. In high dynamics, however, it

becomes a function of the dynamics itself. Scale factors are usually determined by lab calibration of the system; in some cases, field calibration is also used [Salychev, 1998].

Axis misalignment is the result of mounting imperfections of the sensors inside the integrated system. It results in a nonorthogonality of the axes that define the inertial coordinate system. As a result, each axis is affected by the other two axes of the triad. In kinematic applications, the effect of misalignment becomes less significant if frequent manoeuvres occur. In this case, the misalignment error changes in a random manner that permits attributing it to the other random errors. However, if a strapdown system is mounted in an aircraft that flies in a straight line, the misalignment error effect will be systematic.

The previously discussed errors represent systematic effects of the sensors that can be deterministically modeled and consequently removed from the measurements. On the other hand, inertial measurement noise represents the overall uncertainty in the sensor model and effects that are time variable and cannot be deterministically modeled, such as bias instability. These remaining errors are characterized by quasi-systematic and random behaviour [Schwarz, 1986]. They are often modeled as first-order Gauss-Markov or random walk processes with model parameters derived from lab experiments. A stochastic model is used to represent the remaining error effects after removing the systematic sensor errors by deterministic models and lab or field calibrations.

The above various error sources contribute to the behaviour of the errors of the INS. These inertial navigation errors are mainly due to sensor errors, initial navigation errors, vehicle dynamics, and the applied mechanization (see Section 3.2). If the local-level frame is adopted for the mechanization of the inertial navigation system, the INS can be considered to have two horizontal channels and a vertical channel. For each channel, a position error, a velocity error and an attitude error are modeled, namely nine errors for the three channels [Maybeck, 1997]. In a standalone navigation process without any sort of external update, the latter nine error components continually grow with time. The reason for such growth is the fact that navigation parameters (position, velocity, and attitude) are the outcome of integrating incremental linear velocity and angular velocity measurements that are corrupted with time. As a result, the sensor and initial errors are also integrated and contribute to the systematic navigation errors.

Short-term error characteristics of the navigation errors are due to the propagation of the inertial sensors noise in the navigation parameters. Through the integration process, the measurement noise is reduced, resulting in smooth position, velocity, and attitude parameters. For a navigation grade inertial system, the spectrum of the freely navigating inertial system is characterized by the Schuler frequency in the low-frequency band and by white noise in the high-frequency band [Schwarz, 1983]. Nevertheless, for low cost grade inertial sensors, the bounded Schuler behaviour of errors is not guaranteed due to the larger instability of sensor biases as well as the high noise accompanying the sensor output measurements, which increase short-term error effects.

3.1.3 Complementary Nature of INS and GPS

Based on the previous discussion (Sections 3.1.1. and 3.1.2), the error characteristics of both DGPS and INS systems can be summarized. Short-term navigation errors from the INS are relatively small, but they degrade rapidly without bound over time [Mohinder and Angus, 2001]. This is simply because INS non-systematic errors are not predictable and therefore need external aiding. A common trend to describe these non-systematic errors is to use a first-order Gauss-Markov process. On the other hand, DGPS position and velocity errors are not as good as INS over the short term, but they do not degrade with time. More explicitly, a DGPS has long-term stability with a homogeneous accuracy. DGPS measurement errors are mostly of a high frequency nature and can be modeled as white noise in the estimation filter design [Mohamed, 1999].

It is then clear that GPS and INS complement each other. On the one hand, INS needs the long-term stability of GPS to have a handle on its non-predictable output behaviour as well as non-modeled portion of systematic errors. Therefore, the GPS-derived positions and velocities are used as the update measurements for the INS-derived position, velocity and attitude. On the other hand, INS short-term highly accurate positioning and attitude information provides trajectory interpolation between GPS updates. In addition, INS, particularly navigation or tactical grade systems, can be used to interpolate between GPS signal outages and help it recover after loss of lock for short periods. However, this might not be possible with low-cost or MEMS-based INS systems due to the relatively large drift of the non-deterministic part of the sensor biases. This causes a dramatic degradation

of the navigation or positioning accuracy even over short-term DGPS outages [Abdelazim et al., 2004].

3.2 INS Mechanization

INS mechanization is the process of determining the navigation states (position, velocity and attitude) from the raw inertial measurements through solving the differential equations describing the system motion. IMU measurements include three angular rate components provided by the gyroscopes and denoted by the 3x1 vector ω_{ib}^b as well as three linear acceleration components provided by the accelerometers and denoted by the 3x1 vector f^b . The subscript (b) refers to the sensor *body* frame, where all inertial measurements typically take place, while the subscript (i) denotes the inertial frame. This means that the angular velocities ω_{ib}^b of the *body* frame are measured with respect to the *inertial* frame and the superscript (ib) indicates that the inertial measurements are expressed as seen by an observer in the *body* frame.

Mechanization is usually expressed by a set of differential equations and typically performed in the local level frame defined by the local east, north and ellipsoid normal. Equation 3.6 shows the mechanization differential equations in the local level frame [Wong, 1988; El-Sheimy, 2002]:

$$\begin{pmatrix} \dot{r}^\ell \\ \dot{v}^\ell \\ \dot{R}_b^\ell \end{pmatrix} = \begin{pmatrix} D^{-1} v^\ell \\ R_b^\ell f^b - (2\Omega_{ie}^\ell + \Omega_{el}^\ell) v^\ell + g^\ell \\ R_b^\ell (\Omega_{ib}^b - \Omega_{il}^b) \end{pmatrix} \quad (3.6)$$

where

r^ℓ is the position vector in the local level frame including the latitude (ϕ), the longitude (λ), and the ellipsoidal height (h),

v^ℓ is the velocity vector in the local level frame (v_{east} , v_{north} , v_{up}),

R_b^ℓ is the transformation matrix from body to local frame as a function of attitude components,

g^ℓ is the gravity vector in the local level frame,

$\Omega_{ib}^b, \Omega_{il}^b$ are the skew-symmetric matrices of the angular velocity vectors $\omega_{ib}^b, \omega_{il}^b$ respectively, and

D^{-1} is a 3x3 matrix whose non zero elements are functions of the user's latitude ϕ and ellipsoidal height (h).

The solution and numerical implementation of the above differential equation are discussed in more detail in several references [El-Sheimy, 2002; Schwarz and Wei, 2000; Salychev, 1998].

3.3 INS/GPS Kalman Filtering Algorithm

The navigation algorithm integrates the INS mechanization equations (see Equation 3.6) to yield the parameters on the left-hand side, namely the position, velocity, and attitude (PVA) of the vehicle. The algorithm takes into account the Earth's rate of rotation and gravity. The navigation algorithm by itself is seldom useful since the inertial sensor errors (mainly sensors biases) and the fixed-step integration errors will cause the PVA solution to diverge quickly. The navigation algorithm must account for these error sources in order to be able to correct the estimated PVA. The most common estimation algorithm used in integrated INS/GPS is the KF. The KF exploits a powerful synergism between GPS and IMU measurements. This synergism is possible, in part, because the IMU and GPS measurements have very complementary error characteristics as discussed in Section 3.1.3. KF is able to take advantage of these characteristics to provide a common integrated navigation implementation with performance superior to that of either subsystem (GPS or INS) [Jay and Matthew, 1999]. In this integration scheme, the GPS derived positions and velocities are used as the update measurements for the IMU derived PVA. The KF error state vector in this case includes the navigation parameters as well as the accelerometer and gyroscope error states.

3.3.1 Dynamic and Measurement Models

The dynamic error model of the navigation parameters (i.e., PVA) can be determined through the linearization of the mechanization model given by Equation 3.6. A simplified form is then obtained, as in Bar-Itzhack and Berman [1988], by neglecting insignificant terms in the resulting linearized model:

$$\begin{aligned}
\delta \dot{r}^\ell &= D^{-1} \delta v^\ell \\
\delta \dot{v}^\ell &= -(2\Omega_{ie}^\ell + \Omega_{el}^\ell) \times \delta v^\ell - \delta R_b^\ell f^b + R_b^\ell \delta f^b + \delta g^\ell \\
\delta \dot{A}^\ell &= E \delta v^\ell + R_b^\ell \delta \omega^b
\end{aligned} \tag{3.7}$$

where

δr^ℓ is the position error state vector in the local level frame,

δv^ℓ is the velocity error state vector in local the level frame,

δA^ℓ is the attitude error state vector in the local level frame,

δg^ℓ is the error in the computed gravity vector in the local level frame,

δf^b & $\delta \omega^b$ are accelerometers bias and gyro drift vectors in the body frame respectively, and

E is a 3X3 matrix whose non-zero elements are functions of the vehicle's latitude and the Earth's radii of curvatures.

Adding the various inertial sensor residuals to the KF, the standard state vector to be solved for contains fifteen states, which includes the nine navigation parameters, three accelerometer biases, and three gyro drift terms. It is common to model the residual stochastic part due to sensor errors as first order Gauss-Markov processes. Based on the mathematical expressions in Equation 3.7 and the Gauss-Markov representation of inertial sensor errors, the inertial state dynamic error model can be written in the following form:

$$\dot{x} = Fx + Gw \tag{3.8}$$

where

x is the error state vector of inertial navigation whose elements include three position errors, three velocity errors, three attitude errors, three accelerometer biases, and three gyro drift terms,

F is the state transition matrix, a detailed elements expression of the matrix F is provided in Mohinder and Angus [2001],

G is a rectangular matrix whose elements may be time variant,

w is a zero-mean Gaussian white noise vector, and

$G w$ represents the covariance matrix of the applied state model.

The measurement model can be expressed in a generic form as:

$$z = H x + v \quad (3.9)$$

where,

z : is the filter measurements,

x : is the filter states,

H : measurement design matrix,

v : measurement noise.

In the case of INS/GPS error filter formulation, the filter update measurements (z) are calculated by differencing the INS navigation information from that of the GPS [KINGSPAD, 2000]:

$$\begin{aligned} z &= X^{INS} - X^{GPS} \\ &= (X + \delta X + n^{INS}) - (X + n^{GPS}) \\ &= \delta X - e_z \end{aligned} \quad (3.10)$$

where,

X is the navigation information

δX is the INS navigation error state vector

$e_z = n^{INS} - n^{GPS}$ is the measurement model noise.

3.3.2 Kalman Prediction and Update Equations

The KF estimates elements of the state vector (x) by using a form of feedback control. Equations of the KF are divided into two groups of equations; time prediction and measurement update [Mohinder and Angus, 2001]. The time prediction equations are responsible for the forward time transition of the current epoch ($k-1$) states to the next epoch (k) states. Time prediction equations are given by:

$$\hat{x}_k(-) = F_{k,k-1} \hat{x}_{k-1}(+) \quad (3.11)$$

$$P_k(-) = F_{k,k-1} P_{k-1}(+) F_{k,k-1}^T + Q_{k-1} \quad (3.12)$$

where

P is the estimated variance-covariance matrix of inertial states,

Q is the system noise matrix,

F is the transition matrix,

(-) denotes the estimated value after prediction and

(+) denotes the estimated value after updating.

The measurement update equations utilize new measured values into the *a priori* state estimate to obtain an optimized *a posteriori* state estimate. The measurement update equations are given as [Gelb, 1989]:

$$K_k = P_k(-)H_k^T [H_k P_k(-)H_k^T + R_k]^{-1} \quad (3.13)$$

$$\hat{x}_k(+) = \hat{x}_k(-) + K_k(Z_k - H_k \hat{x}_k(-)) \quad (3.14)$$

$$P_k(+) = P_k(-) - K_k H_k^T P_k(-) \quad (3.15)$$

where

K is the Kalman gain matrix,

Z is the vector of updating measurements of position and velocity,

R is the measurements variance-covariance matrix, and

H is the design matrix that relates the measurements to the state vector.

The KF update engine is triggered at every GPS measurement epoch using the difference between the GPS and the INS mechanization solutions as input. Hence, the KF generates an updated estimate for minimizing the INS errors using measurement update equations. Whenever the GPS signal is missing, the KF works in time prediction mode to estimate the error state vector. The block diagram of the integration approach is shown in Figure 3.1. KF equations require the statistical properties of the system to be stationary and well defined which cannot be guaranteed especially in a dynamic environment.

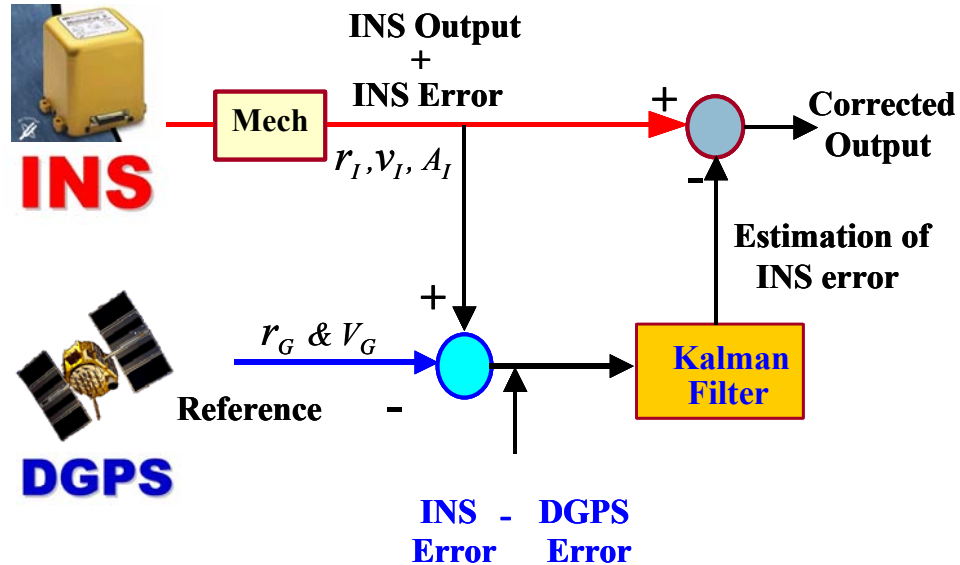


Figure 3.1 Functional block diagram of KF-based IMU/GPS Integration during GPS signal availability

3.4 INS/GPS Integration Schemes

The types of integration can be categorized by the extent to which data from each component aids the function of the other. The first type of integration is coupling of the systems and depends on the mechanization or the architecture of the system. The second type of integration by the method of combining or fusing the data to obtain position coordinates [Phillips and Schmidt, 1996].

There are various architectures for fusing inertial sensors with GPS. Loose integration is the simplest method of coupling. In this scheme, GPS and the inertial sensors generate navigation solutions independently (position, velocity and attitude). The two independent navigation solutions are subsequently combined to form a blended (or filtered) GPS-

inertial navigation solution. One of the benefits of loose integration is that the integrated navigation solution tends to have higher bandwidth and better noise characteristics than the GPS solution alone [Brenner, 1996]. This configuration is best implemented with higher quality inertial sensors (navigation or tactical grade) if the GPS outages are long in duration. Lower quality inertial sensors can also provide some immunity against momentary GPS outages, especially if their outputs were calibrated using GPS prior to the outage. In this instance, the GPS-inertial loose integration is said to include feedback, whereby the difference between the GPS and inertial solutions is fed back to the inertial sensors to carry out the calibration. In general, lower quality inertial sensors (consumer or automotive grade) are suited for applications where GPS outages are infrequent and short in duration [Alban, 2002].

A more complex level of coupling is tight integration, where GPS pseudoranges, Doppler, or carrier phase measurements are blended with the navigation solution generated by the inertial sensors. In addition to the benefits of loose coupling, a tightly integrated system can have a more accurate navigation solution because the basic GPS observables used in the blending process (i.e. pseudoranges and Doppler measurements) are not as correlated as the position and velocity solutions used in loose integration [Farrell and Barth, 1999]. Furthermore, tight integration provides a means for implementing a more sensitive fault detection and isolation scheme that can be used to verify the quality of pseudorange and Doppler measurements [Diesel and Lu, 1995]. Applications which use carrier-phase output (attitude determination and carrier-phase positioning) benefit from tight integration because integer ambiguities can be recovered

and verified quickly from the navigation outputs, despite cycle slips and increased carrier-phase noise [Alban, 2002]. In other words, the real-time feedback of INS velocities to the GPS receiver enables an accurate prediction of GPS pseudorange and phase at the following epoch, thus allowing a smaller bandwidth of the receiver's tracking loop in a high-dynamic environment with subsequent increase in accuracy. The loosely-coupled filtering approach has been highly popular due its modularity and smaller filter size. Although the argument for choosing either form of the implementation has been equal, the tightly-coupled approach is currently gaining more acceptance mainly due to the rapid increase in computational power [Skaloud, 1999].

There are two basic categories of processing algorithms: centralized and de-centralized. In centralized processing, the raw sensor data is combined preferably using one central processor to obtain a position solution. This kind of processing is usually associated with tight system integration. Decentralized processing is a sequential approach to processing, where processors of individual systems provide solutions that subsequently are combined with various degrees of optimality by a master processor. In principle, if the statistics of the errors are correctly propagated, the optimal decentralized and centralized methods should yield identical solutions [Schwarz and Wei, 2000].

In some certain cases, such as system fault detection, isolation, and correction capability, the relative computational simplicity makes the decentralized approach more favourable. The centralized approach provides the best performance in navigation solutions that implement a single robust KF model. The vector state estimation can be implemented in a

closed or open loop, whether or not the estimated sensor errors are fed back to correct the measurements. When properly designed, the closed-loop implementation generally has better performance and is therefore the preferred implementation when using a strapdown INS [Skaloud, 1999]. Figure 3.2 and Figure 3.3 show the architecture of open-loop and closed-loop centralized INS/GPS integration respectively while Figure 3.4 and Figure 3.5 illustrate same but for a decentralized integration. In addition, Table 3.1 presents a comparison of centralized versus decentralized INS/GPS integration including the advantages and disadvantages of each. As a final remark in this section, it is worthwhile to mention that the decentralized INS/GPS integration will be the one adopted herein and will be implemented for processing any integrated inertial and GPS measurements throughout the rest of this thesis.

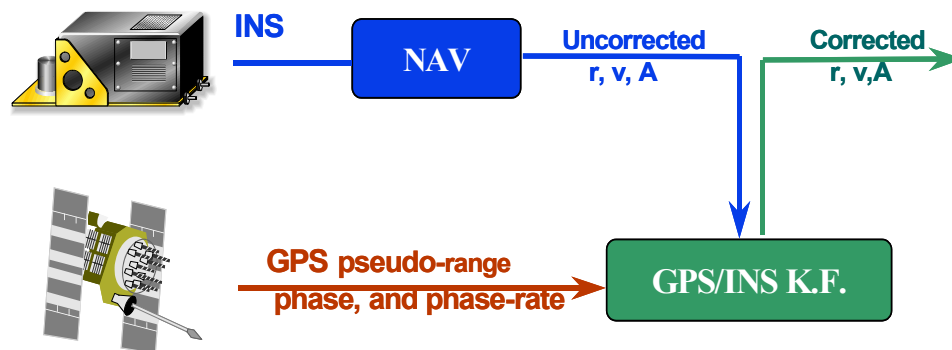


Figure 3.2 Architecture of a centralized INS/GPS integration (open loop) – after [El-Sheimy, 2004]

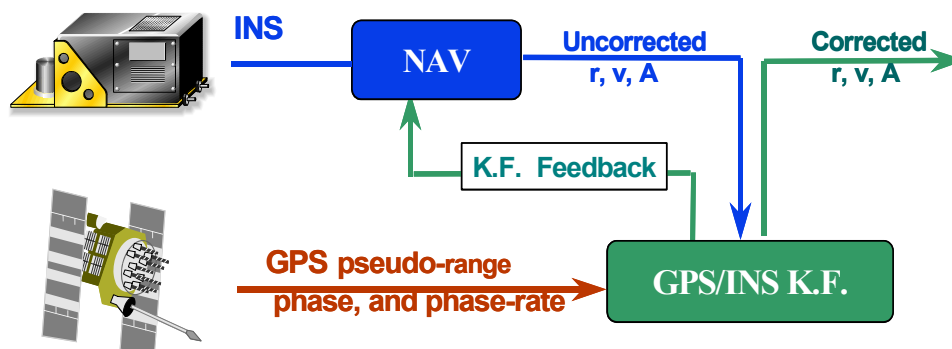


Figure 3.3 Architecture of a centralized INS/GPS integration (closed loop) – after [El-Sheimy, 2004]

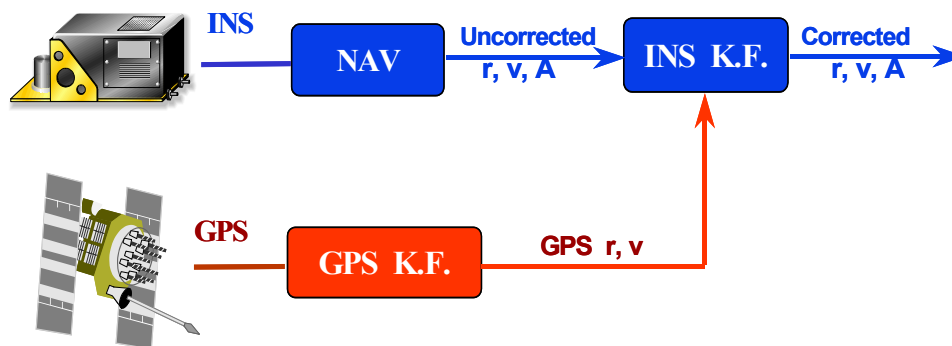


Figure 3.4 Architecture of a decentralized INS/GPS integration (open loop) - after [El-Sheimy, 2004]

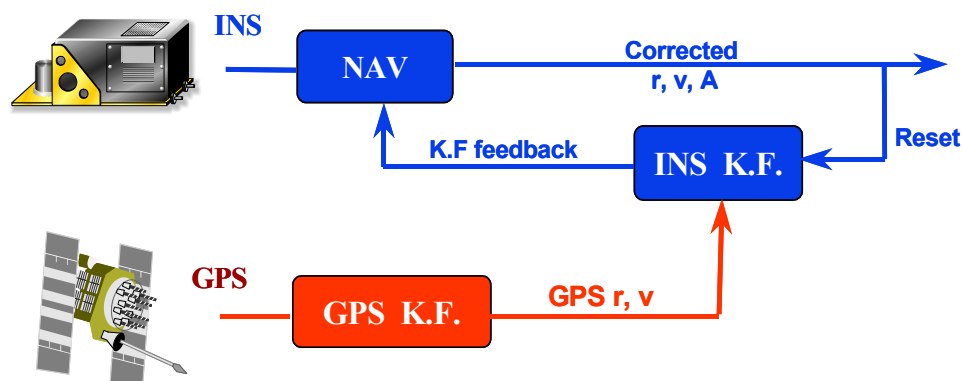


Figure 3.5 Architecture of a decentralized INS/GPS integration (closed loop) – after [El-Sheimy, 2004]

Table 3.1 Decentralized against centralized INS/GPS integration

Decentralized INS/GPS Integration	Centralized INS/GPS Integration
- Consists of GPS receiver, an IMU and integrated filter.	- Consists of GPS receiver, an IMU and integrated filter.
- Uses positions and velocities from GPS every 1-10 seconds.	- Receives range and range rate measurements from each satellite at a 1 Hz rate.
- No solution if less than 4 satellites (far from optimal).	- Optimal use of however many satellites available.
- More susceptible to jamming.	- Better jamming resistance (reduced tracking bandwidth).
- Lower performance in case of high dynamics (receiver may lose lock and take long time to relock on phase).	- Reliable tracking under high dynamics (with the help of inertial measurements, tracking loops quickly relock on phase)
- Simple and flexible approach (can use multi-sensor aid)	- Difficult to apply multi-sensor aid (cannot augment other additional source of information)
- Small filter size	- Larger filter size

3.5 Limitations of INS/GPS Kalman Integration

The integration of INS and GPS measurements utilizing a KF can lead to inertial error compensation within a limited bandwidth that depends on the vehicle dynamics, updating rate and inertial sensor accuracy [Bruton et al., 1999; Chiang et al., 2003]. Inertial sensor errors are classified into two main groups according to their spectral signature *long-term* errors that can possibly be dampened by GPS aiding and *short-term* errors that require a different treatment. For instance, the deterioration in attitude accuracy due to rapid changing errors like system noise and vibrations is not detectable by GPS over a short time period [Bruton et al., 1999]. Considering this concept, short-term inertial noise needs to be suppressed, prior to mechanization and integration, through an optimal pre-filtering tool. The proposed tool, in this thesis, to treat and overcome this shortcoming is the WMRA method. The detailed description, mathematical foundation and experimental validation of WMRA are given later in the following Chapter (Chapter four).

Another limitation is that the KF only works well under certain predefined models and sensor outputs that fit properly with these models. In fact, it is extremely difficult to set an accurate stochastic model for each inertial sensor (gyroscopes and accelerometers) that works efficiently in all cases and reflects the long-term behaviour of these sensors' errors [Hide et al., 2002]. The situation becomes more dramatic if the implemented inertial sensors, in the INS/GPS system, are low-cost or MEMS-based. The current MEMS-based sensors have very high and varying drift rates that are extremely hard to be conveniently described. This will consequently forestall the KF integration algorithm from functioning

properly particularly in INS stand-alone mode when the GPS update measurements are absent. Therefore, the performance of integrated low-cost INS/GPS will dramatically degrade over time during GPS signal outages. This necessitates the augmentation of another effective tool or prediction algorithm that can provide more reliable position and velocity estimates in standalone inertial navigation mode (see Figure 3.6). Fuzzy logic-based networks prediction is proposed in this thesis to handle the latter mentioned KF limitation. Detailed description, mathematical foundation and different types of Fuzzy logic based networks are given in Chapter five. In addition, two different techniques for augmenting one Fuzzy logic algorithm with a conventional Kalman, to solve a MEMS-INS/GPS integration problem, are presented and experimentally verified in Chapter seven.

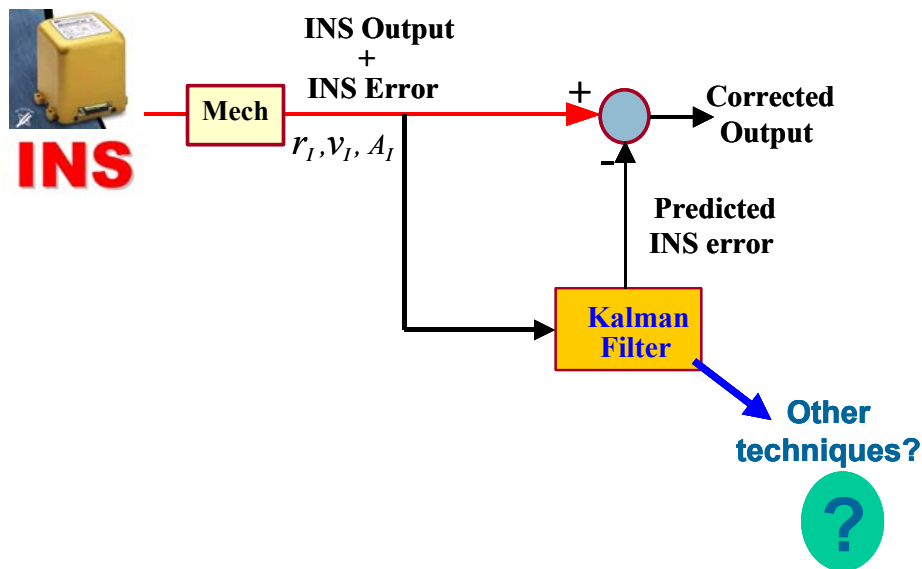


Figure 3.6 Functional block diagram of KF-based IMU/GPS Integration during GPS signal availability

CHAPTER FOUR

WAVELET MULTI-RESOLUTION ANALYSIS OF MEMS INERTIAL SIGNAL

The Wavelet Multi-Resolution Analysis (WMRA) is studied and verified in this chapter as a proposed tool to improve the performance of MEMS-based inertial sensors. A preliminary analysis of MEMS-based accelerometer and gyroscope measurements in the frequency domain is presented first. This provides insight into the distribution of different types of inertial sensor errors prior to applying the wavelet transform (WT). WMRA is then implemented to improve the sensors' signal-to-noise ratios, remove sensor errors that are mixed with motion dynamics, and provide more reliable data to the KF based MEMS-INS/GPS integration module. Applying WMRA to the MEMS inertial signal comprises two main steps. The first involves eliminating the high frequency sensor noise using wavelet de-noising methods. The second step then follows by specifying a proper threshold through which the motion dynamics can be separated from the short-term and/or long-term sensor errors as well as other disturbances.

4.1 Observability of Different Inertial Error Bandwidths

GPS and INS sensors have been historically integrated as the standard to develop a reliable and stable vehicle navigation system. The aim of the navigation process is to fuse the information provided by INS and GPS to accurately estimate the vehicle trajectory. The traditional and common way to perform this INS/GPS integration is to use a KF (see Section 3.3). In the decentralized architecture of KF based INS/GPS integration, two

filters are implemented and performed independently; one for the GPS and the other for the INS. The INS filter is the primary one that uses the raw inertial sensor data to estimate the states (position, velocity and attitude) through the trajectory utilizing mechanization process (see Section 3.2). The output of the GPS filter is then used to update the INS filter to help in the estimation of the INS long-term errors. In order to estimate the INS errors, the INS filter depends on measurements and a dynamic model. The quality of the final estimates of the state depends therefore on the quality of both the measurements being made and the models being used [Titterton and Weston, 1997; Schwarz and Wei, 1990].

The integration of INS and GPS measurements can lead to inertial error compensation within a limited bandwidth that depends on the vehicle dynamics, updating rate and inertial sensor accuracy [Bruton et al., 1999]. Inertial sensor errors are classified into two categories according to their spectral signature: *long-term errors* that can possibly be dampened by GPS aiding and *short-term errors* that require a different treatment. For instance, the accuracy deterioration due to rapidly changing errors like system noise and engine vibrations is not detectable by GPS over a short time period [Skaloud et al., 1999].

Considering the above concept, better INS/GPS system performance can be expected if short-term inertial noise is suppressed prior to mechanization and integration. Hence, optimal low-pass filtering beyond the motion bandwidth should be applied before inertial data mechanization. In addition, it can be shown from Figure 4.1 that these short-term errors are mixed with the motion dynamics. Therefore, and particularly in kinematic

applications, pre-filtering has to be extended further to frequency bands where noise and vibrations overlap with the signal of interest in a frequency range at which external aiding is not capable of separating them [Skaloud et al., 1999].

Long-term errors are reduced by updating the filter with the error state vector that is generated from the GPS filter (position and velocity). The short-term errors are partially reduced by the smoothing which results from the numerical integration process of the INS mechanization (see Figure 4.1). Pre-filtering the inertial sensor measurements utilizing WMRA to separate the sensor errors from the motion dynamics can also reduce most of the remaining effects of the short-term errors.

Based on an error model for the different navigation system parameters (including stochastic models for sensor errors), a KF solution is capable of providing a reliable estimate of the position, velocity, and attitude components of the moving platform [Schwarz & Wei, 1990]. In addition, KF can predict the vehicle position during GPS signal blockage relying on the localized INS outputs and the pre-designed error model.

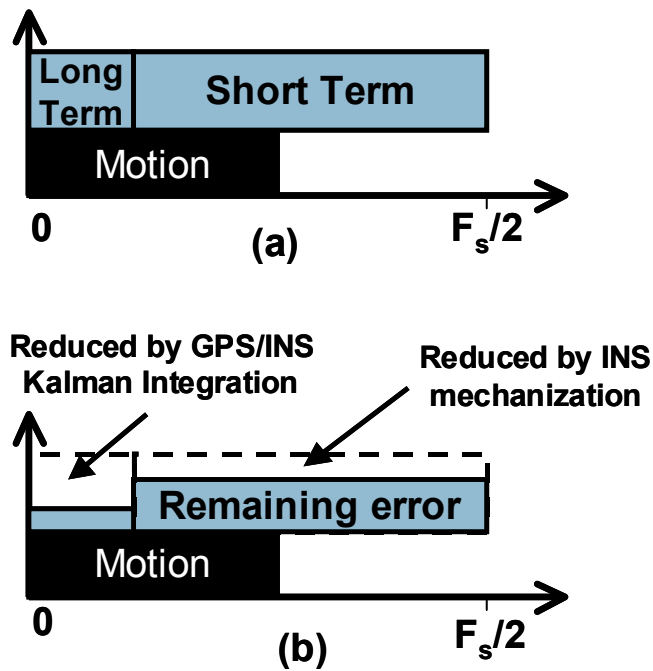


Figure 4.1 Error bandwidth in an INS system: a) raw measurements, b) after standard processing (F_s is the sampling frequency) [after Skaloud et al., 1999]

Driven by their low cost and small size, MEMS sensors have been used to produce low cost INS that can be widely adopted in several navigation applications. However, KF based INS/GPS integration techniques may not be suitable for MEMS-based navigation systems. This is due to two main reasons (see Section 3.5). The first reason is that KF suppresses the effect of inertial sensors' noise using GPS position and velocity update but within a limited band of frequency. If the same technique is applied to MEMS-based INS, the overall accuracy may be jeopardized due to the significance of those MEMS-based inertial sensor errors that are mixed with the motion dynamics. The second reason is that a KF needs a certain predefined and accurate stochastic model for each inertial sensor (gyroscopes and accelerometers) that works efficiently in all cases and reflects the long-

term behaviour of the corresponding sensor error. In fact, this is extremely difficult to achieve if the MEMS-based inertial sensors are involved in the integrated INS/GPS system due to their relatively high measurement noise. Therefore, it is essential to reduce the noise level at the output of MEMS-based inertial sensors and separate the sensor errors from motion dynamics prior to processing their measurements by the KF module.

4.2 Power Spectrum In Static and Kinematic Modes

Linear transforms, especially Fourier transform (FT), are widely used in solving problems in science and engineering. The FT, as a pervasive and versatile tool, is used in many fields of science to alter a problem into one that can be more easily solved [Bracewell, 1965]. Specifically, the FT is used in linear systems analysis, random process modeling, probability theory, quantum physics, and boundary-value problems [Brault and White, 1971]. The FT, in essence, converts a waveform or function into sinusoids of different frequencies, which sum to the original waveform. It identifies or distinguishes the different frequency sinusoids and their respective amplitudes [Brigham, 1988]. The discrete Fourier transform (DFT) $X(\omega)$ of a discrete time signal $x(n)$ is given by:

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n) \cdot e^{-j\omega n} \quad (4.1)$$

Analytic and graphical representation of the power spectrum function is of great value and it will be quite beneficial if applied to inertial sensors. The careful interpretation of this spectrum helps in recognizing the frequency band of all dynamics existing in the signal including the signal of interest representing the actual motion of the vehicle as well as the disturbing noise components contaminating the original signal. This disturbance

can be classified mainly into sensor noise which is normally Gaussian, and secondary dynamics being coloured or having a sinusoidal signature. The latter noise type usually arises in kinematic mode due to vehicle motion but it cannot be considered as part of the motion of interest. In order to separate these sinusoidal noise components and remove them from the true motion signal we have to identify their frequency band and consequently select the proper filter. A simple and common way to do this is to analyze the power spectrum of inertial sensor output signals in both static and kinematic modes. Figure 4.2 to Figure 4.5 show the power spectral function of two MEMS inertial sensors (an accelerometer and a gyroscope) taken from a real navigation field test where the raw output data rate of both sensors is 200 Hz. Figure 4.2 and Figure 4.3 depict the lateral acceleration as a typical example of linear components while Figure 4.4 and Figure 4.5 are related to the pitch rate as a typical example of angular inertial measurements.

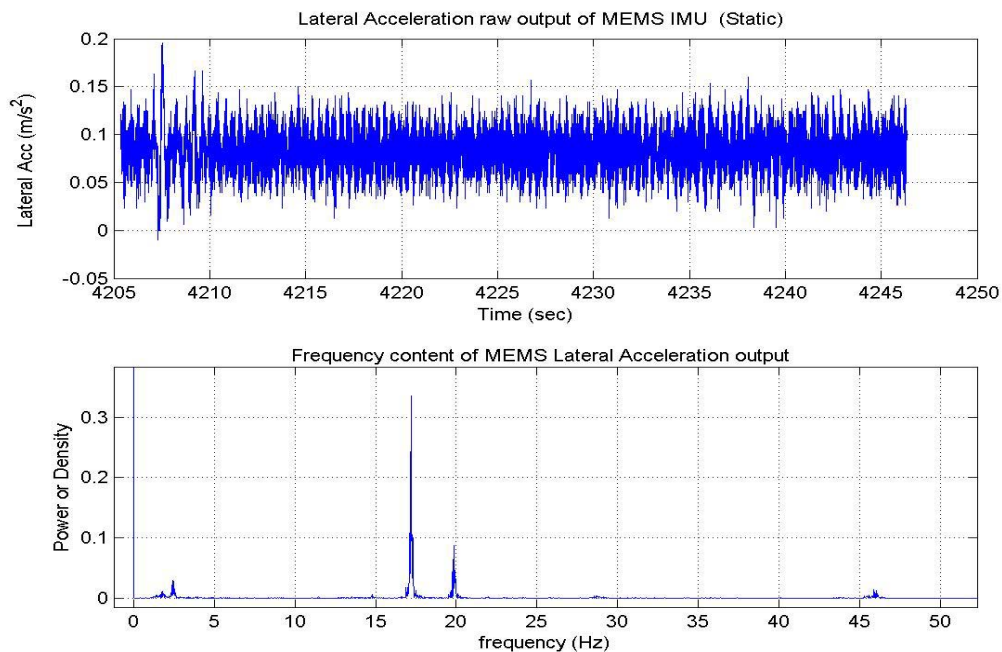


Figure 4.2 Power spectrum of MEMS lateral acceleration in static mode

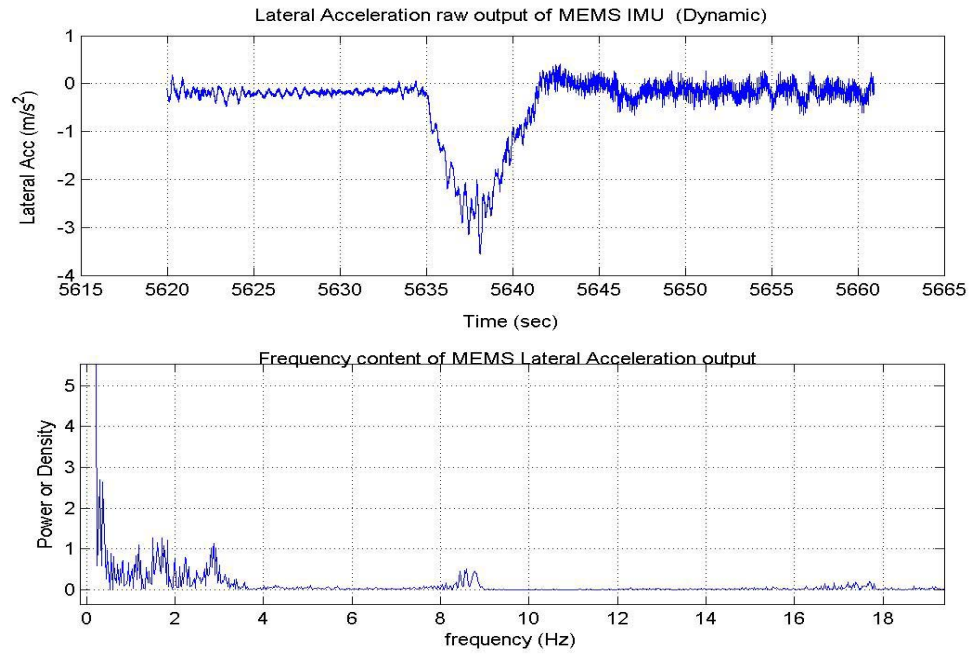


Figure 4.3 Power spectrum of MEMS lateral acceleration in kinematic mode

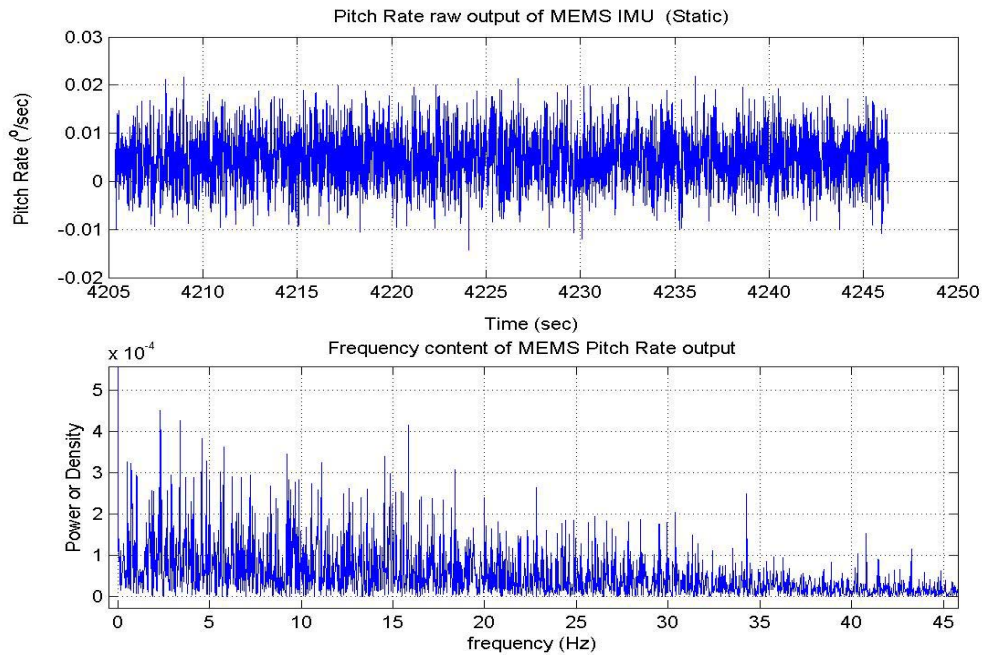


Figure 4.4 Power spectrum of MEMS pitch rate in static mode

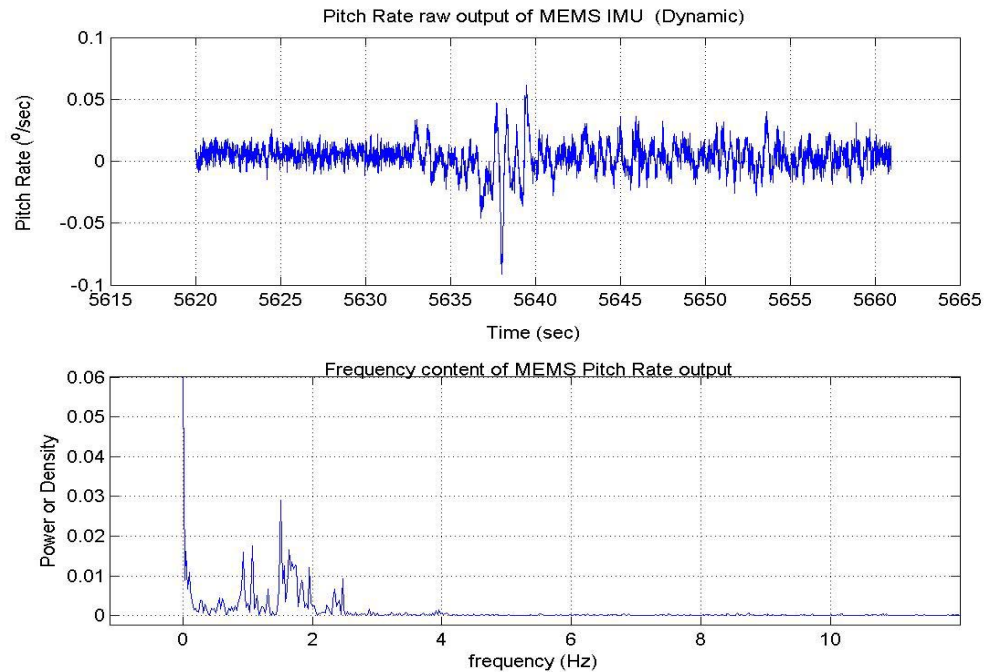


Figure 4.5 Power spectrum of MEMS pitch rate in kinematic mode

It can be deduced from the above figures that the sensor Gaussian noise always exists either in static or in dynamic mode and it is distributed throughout the whole observed frequency range. In addition, the useful signal pertaining to the motion of interest has almost zero frequency in static mode for both signals. In dynamic mode, the monitored vehicle motion frequencies exist in a very low frequency band close to zero for pitch rate and a relatively wider frequency band (0 to 3Hz) for lateral acceleration. It is also clear that an interfering sinusoidal noise for both signals, only in dynamic mode, appears within a frequency range between 1Hz and 3Hz that overlaps the useful signal bandwidth. The source of this interfering signal might be secondary dynamics such as pumping, or

tire and pavement irregularities. Some vibrations, most probably caused by the vehicle engine, exist between 8 and 9 Hz away from that of the motion frequencies.

4.3 Wavelet Multi-Resolution Analysis and Thresholding

For real-time application, the difficulty in analyzing a non-stationary signal is related to determining the moment in time at which a given transient occurred. Time localization is essential to capture the non-repetitive frequencies in a non-stationary signal [Ogden, 1997]. It has been shown that better time localization can be accomplished by applying short time FT (STFT), which utilizes a window function that is multiplied by the input signal before computing the FT [Robertson et al., 1996]. Although STFT provides a time-frequency representation of a signal, there is a major drawback with respect to utilizing STFT in processing inertial sensor data; namely that the width of the window is fixed. The STFT of a signal can be represented in a two dimensional grid as shown in Figure 4.6 where the divisions in the horizontal direction represent the time interval for each window; the divisions in the vertical direction represent the frequencies; and the shade of each rectangle is proportional to the frequency of the monitored signal component. The darker the shade of the rectangle, the lower the frequency of the signal component that can be observed. Obviously, there is a time-frequency trade-off corresponding to the choice of the window width. If Figure 4.6 (a), (b) and (c) are examined and compared to each other, it can be seen that, as the width of the window function decreases, a smaller portion of the input signal is considered. This means that, when a large window is used (see Figure 4.6.a), the low frequency components of the signal are observed while when a small window width is used (see Figure 4.6.c), the high frequency components are mainly

detected. As the width of the window function increases, more accurate information about the different frequencies within the window are obtained, but the ability to determine when these frequencies occur is lost [Ogden, 1997]. Thus there remains a need for multiple resolutions that can provide fine time resolution for short duration and high frequency signals, and fine frequency resolution for long duration and lower frequency signals [Strang and Nguyen, 1997].

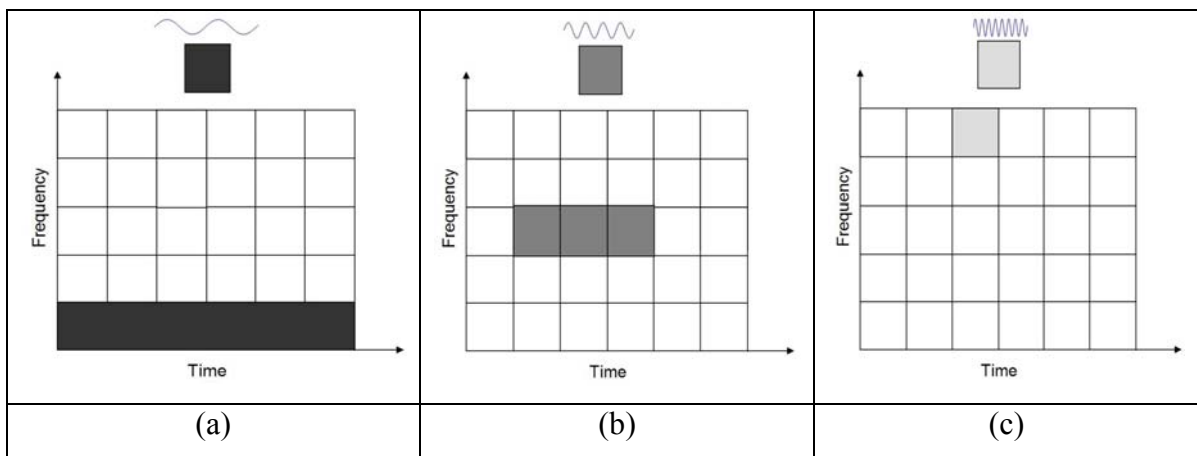


Figure 4.6 A scalogram illustrating Short-time Fourier Transform [after Robertson et al., 1996].

4.3.1 Continuous and Discrete Wavelet Transform

An efficient digital signal-processing algorithm needs to provide multiple resolutions in time and frequency domains. More precisely, such an algorithm should provide fine time resolution for short durations and high frequency components of the signal as well as for long durations and low frequency parts of the signal. In addition, this algorithm should be

capable of providing accurate location of the transient component while simultaneously retaining information about the fundamental frequency and its low-order harmonics. Wavelet analysis, therefore, represents the next logical step, which is based on a windowing technique with variable-sized regions. In fact, wavelet analysis is capable of utilizing long time intervals where precise low frequency information is needed, and shorter intervals where high frequency information is considered. In general, the major advantage offered by wavelets is the ability to perform local analysis; that is to analyze a localized area of a larger signal [Mallat, 1989]. Therefore, WT has been recently adopted for INS to monitor the acceleration and angular velocity acquired by the accelerometers and the gyroscopes mounted inside the vehicle.

Continuous wavelet transform (CWT) of a time domain signal $x(t)$ is obtained using the following equation [Strang and Nguyen, 1997]:

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-b}{a}\right) dt \quad (4.2)$$

where a and b are, respectively, the scaling and shift (position) parameters of the wavelet function $\psi(t)$. Each scale a and position b result in a particular scaled and shifted version of the wavelet function that is to be multiplied by the time domain signal $x(t)$. The integration over time give the wavelet coefficient that corresponds to the scale a and the position b .

The basis function $\psi(t)$ is not limited to exponential (or sinusoidal) base functions like the case of FT or STFT. The only restriction on $\psi(t)$ is that it must be short and

oscillatory. Furthermore, it must have zero average and decay quickly at both ends. These restrictions ensure that the integration in the WT of Equation 4.2 is finite. The function $\psi(t)$ has been given the name wavelet or “small wave” and is referred to as the “mother wavelet” and its dilates (scaled) and translates (shifted) simply as “wavelets” or “daughters wavelets” [Mallat, 1989]. Obviously, the WT of a time-domain signal is defined in terms of the projections of this signal on to a family of functions that are all normalized dilations and translations of a wavelet function.

The discrete wavelet transform (DWT) is utilized in this study for the use with digital signal processing modules. The DWT of a discrete time sequence $x(n)$ is given as [Walker, 1999]:

$$C_{j,k} = 2^{(-j/2)} \sum_n x(n) \psi(2^{-j} - k) \quad (4.3)$$

where $\psi(n)$ is the wavelet function (the basis function utilized in the wavelet transform) and $2^{(-j/2)} \psi(2^{-j} - k)$ are scaled and shifted versions of $\psi(n)$ based on the values of j (scaling coefficient) and k (shifting coefficient) and is usually written as $\psi_{j,k}(n)$. The j and k coefficients take integer values for different scaling and shifted versions of $\psi(n)$. $C_{j,k}$ represents the corresponding wavelet coefficients. Therefore, the original signal $x(n)$ can be generated from the corresponding wavelet function using Equation 4.3:

$$x(n) = \sum_j \sum_k C_{j,k} \psi_{j,k}(n) \quad (4.4)$$

As mentioned earlier, the basis functions $\psi_{j,k}(n)$ are not limited to exponential functions as in the case of FT and STFT. In order to guarantee that the above summation gives a finite value, the only restriction on $\psi_{j,k}(n)$ is that it must be short and oscillatory (it must have zero average and decay quickly at both ends). Figure 4.7 shows two examples of mother wavelets (Morlet and Mexican Hat Wavelets) demonstrating their oscillatory and potentially non-sinusoidal nature [Daubechies, 1988].

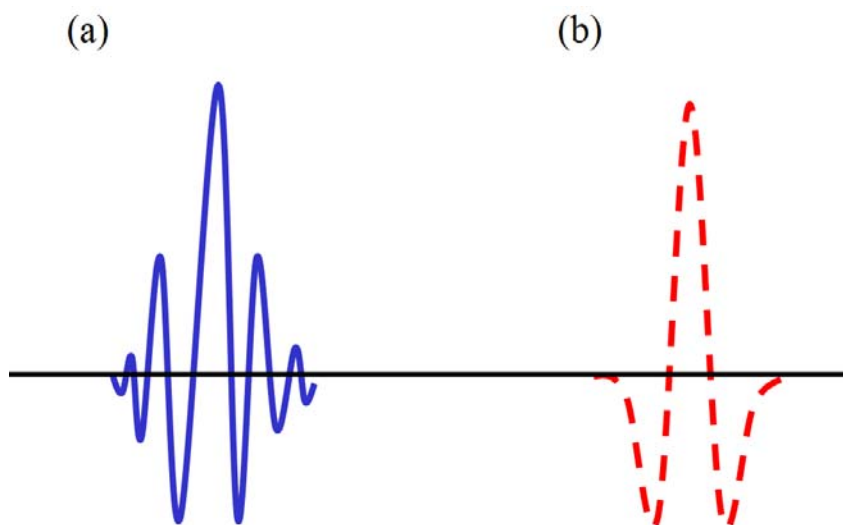


Figure 4.7 Example of wavelet functions.

The DWT output can be represented in a two-dimensional grid similar to the STFT but with very different divisions in time and frequency as shown in Figure 4.8. The rectangles in Figure 4.8 have equal areas or constant time-bandwidth products so that they narrow at the low scales (high frequencies) and widen at the high scales (low frequencies) [Chui, 1992]. In contrast to the STFT, the DWT isolates the transient high

frequency components in the top frequency band precisely at the time of their occurrence while the continuous low components are presented as a continuous magnitude. This illustrates how the multi-resolution properties of the DWT are well suited to transient signals superimposed on continuous fundamental signals.

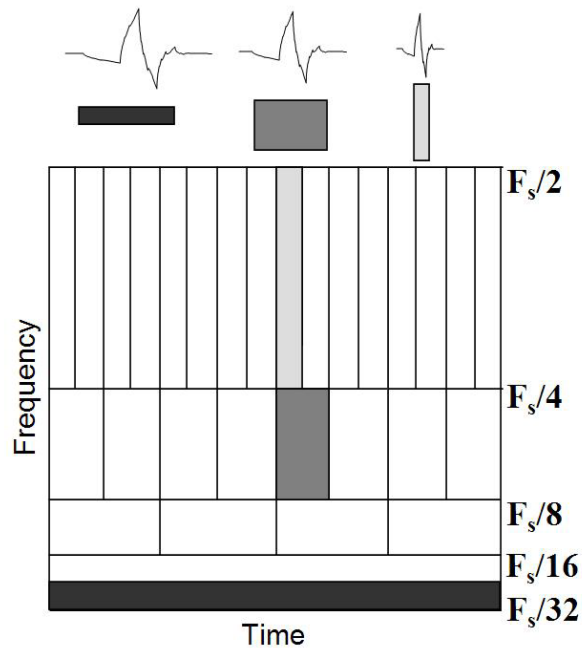


Figure 4.8 A scalogram representing Discrete Wavelet Transform [after Robertson et al., 1996].

4.3.2 Wavelet Multi-Resolution Analysis (WMRA)

WMRA is a technique used to perform the DWT. It allows the decomposition of signals into various resolution levels. The data with coarse resolution contains information about low frequency components and retains the main features of the original signal. The data with fine resolution retains information about the high frequency components.

Scaling a wavelet simply means stretching or compressing it in the time domain. The smaller the scale the more the wavelet will be compressed while the larger the scale the more the wavelet will be stretched. Therefore, low scales allow the analysis of rapidly changing details (high frequency components) and high scales allow the analysis of slowly changing features (low frequency components) [Burrus et al., 1998]. The low frequency component of the signal is usually the most important part that identifies the long-term variation of the signal and is capable of providing a very good *approximation* about it [Burrus et al., 1998]. The approximations correspond to the high scale low frequency part. On the other hand, the high frequency content carries few *details* about the signal. The details correspond to the low scale high frequency part. WMRA therefore decomposes the signal into various resolution levels. The data with coarse resolution contains information about low frequency components and retains the main features of the original signal. The data with fine resolution retains information about the high frequency components [Chui, 1992].

Consider j and k to be the dilation (scaling) index and the translation (shifting) index, respectively. Each value of j corresponds to analyzing a different resolution level of the signal. The mathematical procedure of WMRA is as follows [Strang and Nguyen, 1997; Chui, 1992; Burrus et al., 1998]:

1. For an input signal $x(n)$, the approximation coefficient $a_{j,k}$ at the j^{th} resolution can be computed as follows:

$$a_{j,k} = 2^{(-j/2)} \sum_n x(n) \phi(2^{-j}n - k) \quad (4.5)$$

where $\phi(n)$ is called the scaling function. Scaling functions are similar to wavelet functions except that they have only positive values. They are designed to smooth the input signal thus operating in a manner equivalent to low pass filtering which rejects high frequency components of the signal [Burrus et al., 1998].

2. The approximation of $x(n)$ at the j^{th} resolution level is then computed as:

$$x_j(n) = \sum_{k=-\infty}^{\infty} a_{j,k} \phi_{j,k}(n) \quad (4.6)$$

3. The details coefficient $d_{j,k}$ at the j^{th} resolution level and the detail signal $g_j(n)$ are then computed as:

$$d_{j,k} = \sum_n x(n) \psi_{j,k}(n) \quad (4.7)$$

$$g_j = \sum_{k=-\infty}^{\infty} d_{j,k} \cdot \psi_{j,k}(n) \quad (4.8)$$

where $\psi_{j,k}(n)$ is the wavelet basis function.

The above three steps are repeated for the $j+1$ resolution level but using the approximation $x_j(n)$ obtained in step 2. The original signal $x(n)$ can be reconstructed using all the details obtained during the decomposition process at all resolution levels as:

$$x(n) = \sum_{j=-\infty}^{\infty} g_j(n) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \cdot \psi_{j,k}(n) \quad (4.9)$$

The above equation implies that one has to process the original signal at an infinite number of resolutions, which is impractical. Alternatively, the analysis can stop at the J^{th}

resolution level and the signal can be reconstructed using the approximation at that level and all the details starting from the first resolution level until the J^{th} level. The following equation presents this procedure [Strang and Nguyen, 1997]:

$$x(n) = \sum_{k=-\infty}^{\infty} a_{j,k} \cdot \phi_{j,k}(n) + \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \cdot \psi_{j,k}(n) \quad (4.10)$$

The first term represents the approximation at level J and the second term represents the details at resolution level J and lower. Therefore, WMRA builds a pyramidal structure that requires a continuous and iterative application of the scaling and the wavelet functions, respectively. These filters initially act on the entire signal band at the high frequency first and gradually reduce the signal band at each stage. As presented in Figure 4.9, the high frequency band outputs are taken as the detail coefficients ($D1, D2, D3$), and the low frequency band outputs are taken as the approximation coefficients ($A1, A2, A3$). At this stage, it is worth mentioning that a down-sampling process is performed at every decomposition stage. If the decomposition operation is performed on the entire range of the signal, the number of samples considered in the analysis will be doubled. For instance, if the original signal consists of 100 samples, the approximation and the detail signals will each have 100 samples (total of 200 samples). In order to overcome this problem, which affects computation time and data storage, the down-sampling process is carried out in which the second sample of each signal is ignored.

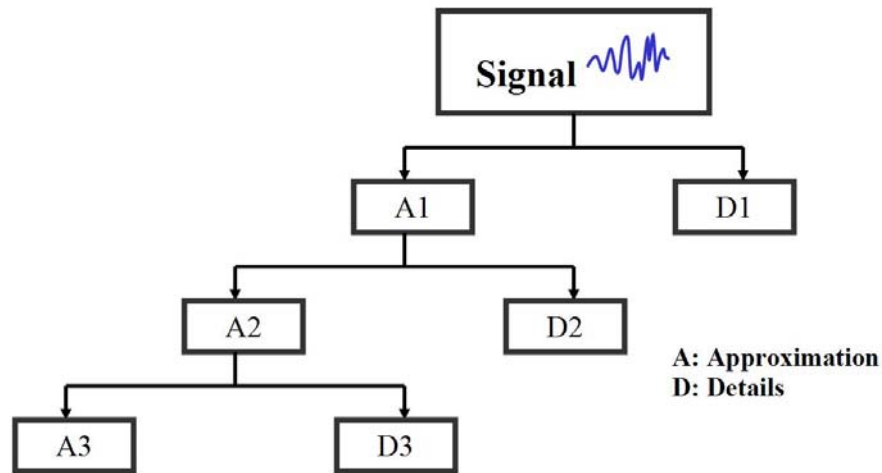


Figure 4.9 Wavelet multi-resolution analysis considering three levels of decomposition.

4.3.3 Wavelet Thresholding

Apart from the Gaussian noise, a sensor signal is often contaminated with sinusoidal interference of a single frequency and its harmonics for some reasons such as inadequate shielding of the sensor electronics or vibrations and secondary dynamics in kinematic systems. Since sinusoidal interference frequency is always within the frequency band of the motion dynamics, a special treatment is required for its separation without degrading the signal of interest (motion dynamics). More explicitly, the successful removal of this sinusoidal interference depends on the effective separation of the main profile of the signal from the noisy components and the accurate estimation of the threshold. Once sinusoidal interference appears in the observations of a sensor signal, subsequent data processing and analysis can lead to erroneous results [Birlasekaran & Ming, 2002]. In fact, the sinusoidal interference can be regarded as coloured ‘noise’ and its transform

coefficients spread into a limited number of sub-bands whereas the transform coefficients of Gaussian noise spread across all sub-bands.

A wavelet thresholding technique is implemented in this study due to its effectiveness for real-time operations. As the detail coefficients may contain sharp transition details of the monitored dynamics, a thresholding operator and threshold estimator must be discreetly selected to ensure effective cancellation of the interference but with a minimum distortion to the sharp transition details of the true signal. Two thresholding operators, the soft and the hard operators, are proposed by Donoho et al. [1994, 1995]. In the case of hard-thresholding, any wavelet coefficient with an absolute value below the threshold is replaced by zero as shown in Figure 4.10.a. Coefficients with an absolute value above the threshold are kept the same. In the case of soft thresholding, coefficients with a magnitude above the threshold are replaced (and hence reduced in value) by the threshold value (see Figure 4.10.b). There is also a more sophisticated shrinking function, which utilizes more complex algorithms to estimate the shrink value. The hard thresholding can be described with the following equation:

$$th(d) = \begin{cases} y & \text{if } |y| > T \\ 0 & \text{if } |y| \leq T \end{cases} \quad (4.11)$$

On the other hand, the soft thresholding operator can be described as:

$$th(d) = \begin{cases} y - T \cdot \text{sign}(d) & \text{if } |y| > T \\ 0 & \text{if } |y| \leq T \end{cases} \quad (4.12)$$

where T is the threshold value and $\text{sign}()$ represents the sign operator. The value of T determines the ultimate effect of the removal of harmonic interference.

Donoho and Johnstone [1994] proposed that T should be selected according to the standard deviation of the Gaussian noise (σ) and the length of the observations (N) as:

$$T = \sigma \sqrt{2 \log N} \quad (4.13)$$

If the noise is white, its standard deviation can be estimated from the median of its finest scale wavelet coefficients, provided that the signal $x(t)$ is piecewise smooth, as:

$$\sigma \approx \frac{1}{0.6745} \text{Med}(|x(t)|) \quad (4.14)$$

where $\text{Med}(|x(t)|)$ is the median value of the sequence $x(t)$.

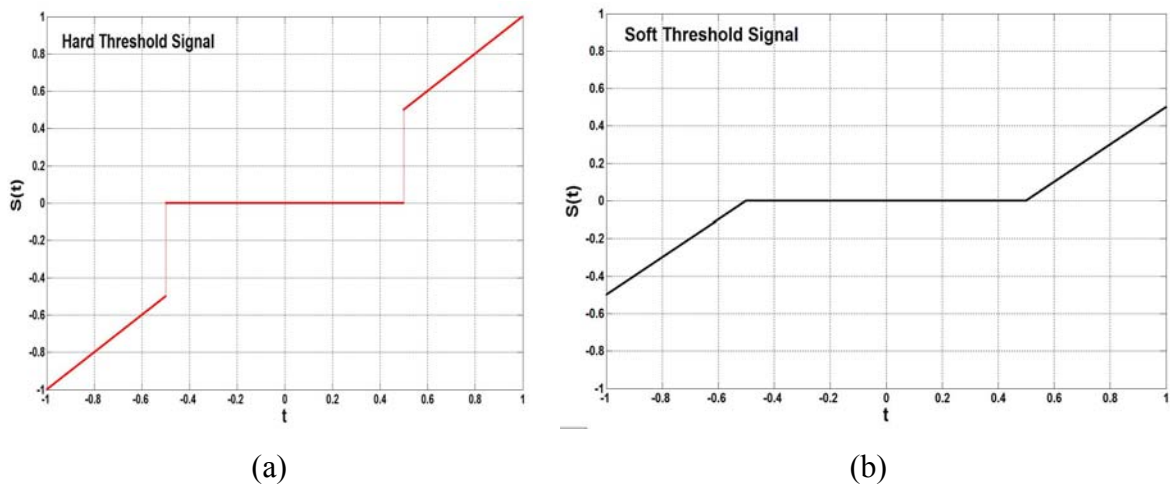


Figure 4.10 Two different threshold operators (a) hard threshold (b) soft threshold .

In case the noisy components are not white, it is recommended to estimate the standard deviation level by level. Since the length of the detail coefficients decreases with the level in the orthogonal wavelet decomposition, Equation 4.7 can be re-written as follows:

$$T = \lambda_j \sigma_j \sqrt{2 \log(N/2^j)} \quad (4.15)$$

where (j) is the decomposition level and λ_j is a level dependent relaxation factor. λ_j serves to achieve a balance between the cancellation of noisy components and distortion to the signal details [Donoho, 1995].

4.4 Wavelet Decomposition of MEMS Inertial Output Signal

The above wavelet thresholding methodology was applied on MEMS inertial sensors mounted inside a land vehicle among other navigation equipment. The ability of the proposed thresholding technique to remove the effect of noise and other short-term errors without deteriorating the true signal of interest is investigated. The output data rate of both linear and angular MEMS inertial sensors was 200 Hz. After applying wavelet analysis to all inertial sensors, their measurements were processed by the KF based INS/GPS integration algorithm.

The Daubechies “db8” wavelet was applied first with hard then with soft thresholding criterion. Figure 4.11 compares the raw measurements of the forward accelerometer to the same measurements after applying the proposed WMRA technique while Figure 4.12 shows the pitch rate before and after processing by WMRA. In both figures, it can be seen that the signal was efficiently de-noised, while all sharp transition details of the true signal remain. It is worthwhile to mention that the resultant denoised accelerations and angular rates after applying soft thresholding were similar to a great extent to those after hard threshold implementation. As a result, the difference between hard and soft

thresholding results is too slight to be observed graphically. This is typically the case in most urban land-vehicle navigation applications where the GPS signal is frequently susceptible to blockage, since sharp transitions in the output inertial signal due to real vehicle motion are rare.

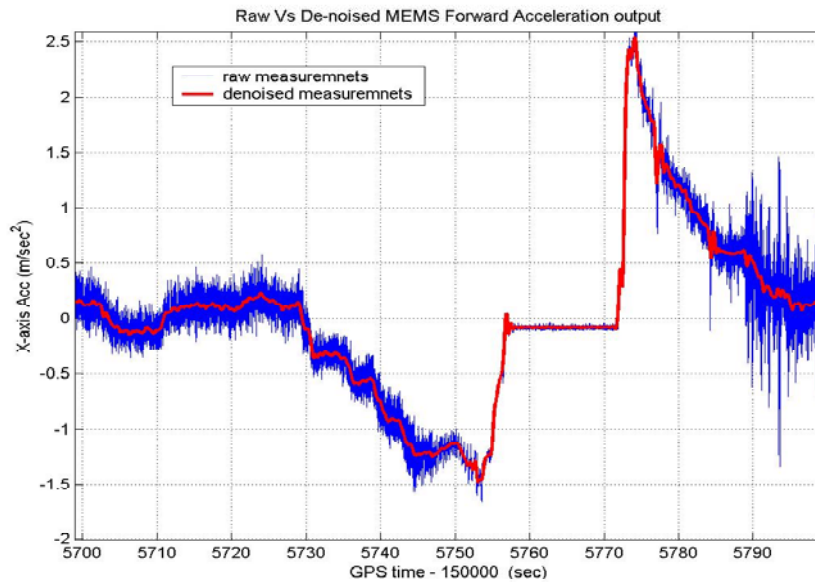


Figure 4.11 MEMS forward acceleration signal before and after wavelet thresholding

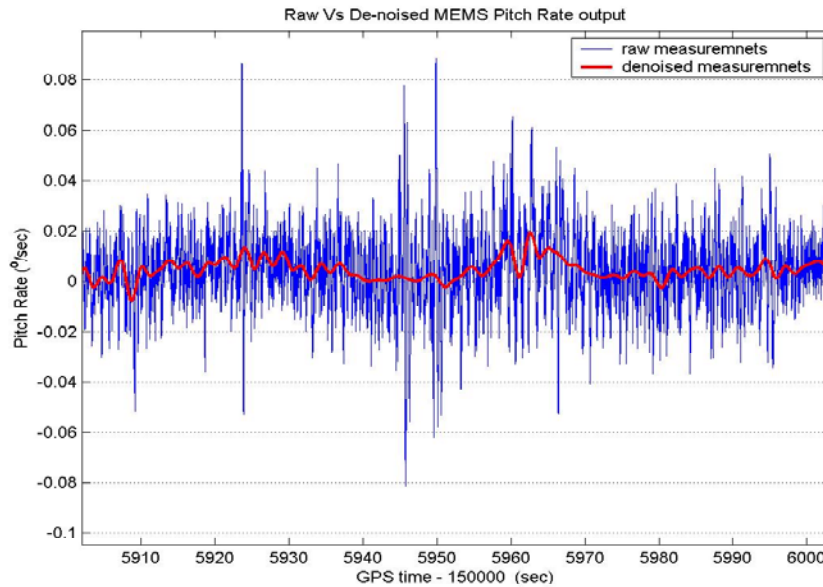


Figure 4.12 MEMS pitch rate signal before and after wavelet thresholding

4.5 Experimental Validation of the Proposed WMRA

In order to examine the effectiveness of the proposed WMRA method, with hard or soft thresholding, and its impact on the accuracy of the computed navigation parameters (particularly position), field test data incorporating MEMS-based INS and DGPS were considered. The field test data were recorded during a van test involving ISI MEMS-based IMU containing *Analog Devices* MEMS gyros and *Colibrys* MEMS accelerometers (see

Table 7.1 for the technical specifications and resolution of the ISI IMU). In addition, the *POS-LV-320* system (Applanix Corporation, Canada), which utilizes a tactical grade *LTN200* inertial measurement unit and a *NovAtel Millennium* GPS receiver, was also involved. The test was performed in an open sky where range measurements from at least

seven to eight satellites were available all the time. A double differenced DGPS solution was performed using another *NovAtel Millennium* receiver at the reference station. The average standard deviations (1σ) of DGPS measurements, in kinematic mode, were (1-2 cm) for horizontal positions, (3-4 cm) for vertical positions, (0.2-0.25 m/s) for horizontal velocities, and (0.3-0.4 m/s) for vertical velocities. Both *DGPS* and *LTN200* trajectories were available throughout the whole test so that they can be used as an accurate reference for the low cost MEMS IMU. A number of intentionally introduced GPS outages of 30 seconds were applied in different situations along the trajectory (see Figure 4.13) to consider different dynamics and motion types.

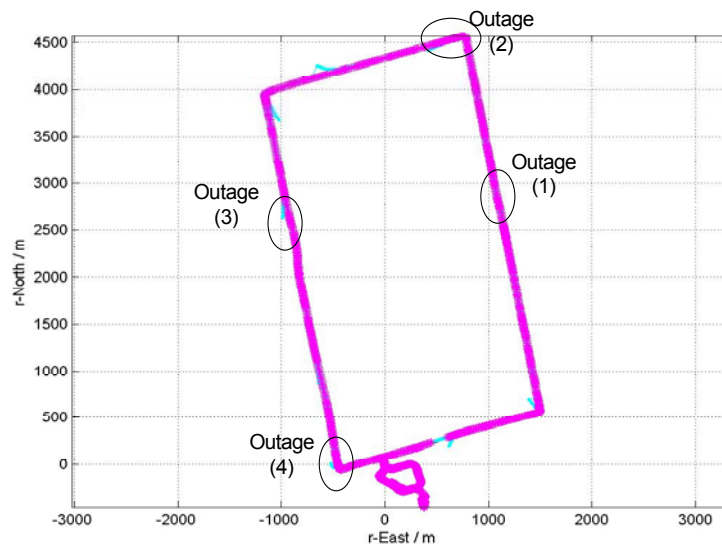


Figure 4.13 Test field trajectory with the simulated GPS outage periods indicated

Figure 4.14, Figure 4.16, Figure 4.18 and Figure 4.19 demonstrate the three position errors ($\delta\phi, \delta\lambda, \delta h$) during four different GPS outages as provided by KF before and after applying the proposed WMRA procedure with soft thresholding. In all these figures, the

horizontal axis represents (GPS time – 150000) while the vertical axis represents the longitude, latitude or altitude error. Vehicle motion was almost straight during two of the presented outages (1 & 3) while there was a ninety-degree (90°) left turn included during the other two outages (2 & 4). Figure 4.14 and Figure 4.15 as well as Figure 4.16 and Figure 4.17 represent a comparison of the position accuracy, during GPS outages, when applying soft thresholding versus that with the utilization of hard thresholding.

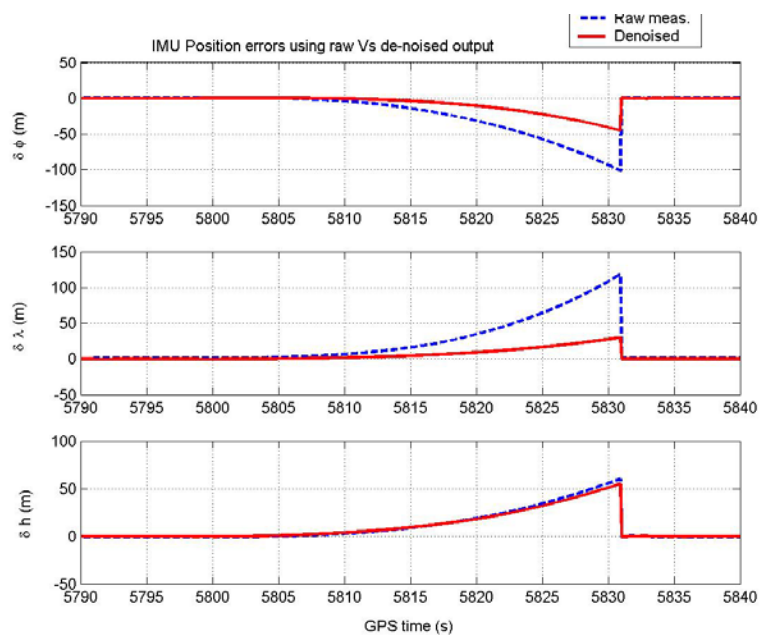


Figure 4.14 Position accuracy during GPS outage (1) – Raw versus de-noised INS measurements [denoising with soft threshold WMRA]

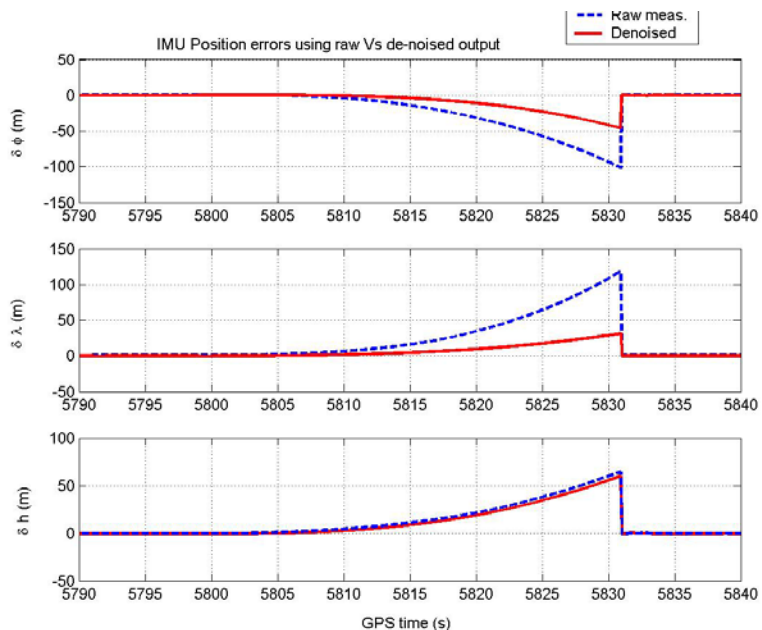


Figure 4.15 Position accuracy during GPS outage (1) – Raw versus de-noised INS measurements [denoising with hard threshold WMRA]

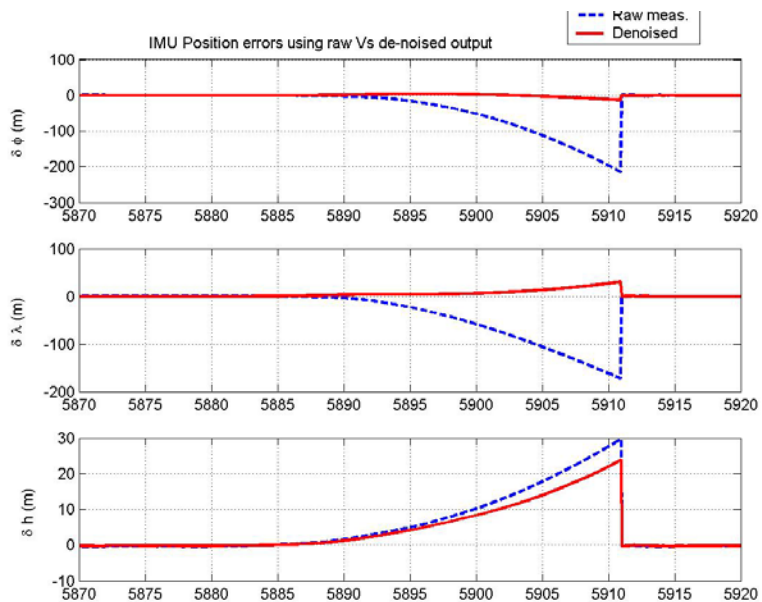


Figure 4.16 Position accuracy during GPS outage (2) – Raw versus de-noised INS measurements [denoising with soft threshold WMRA]

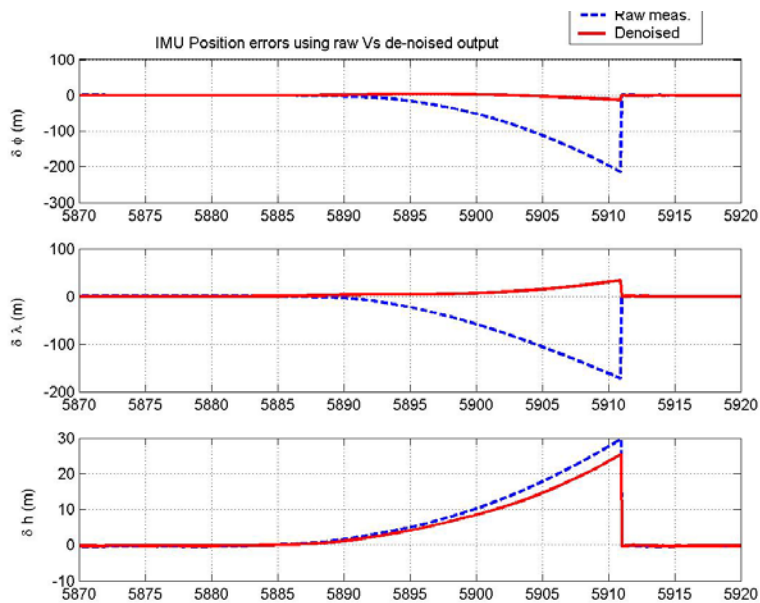


Figure 4.17 Position accuracy during GPS outage (2) – Raw versus de-noised INS measurements [denoising with hard threshold WMRA]

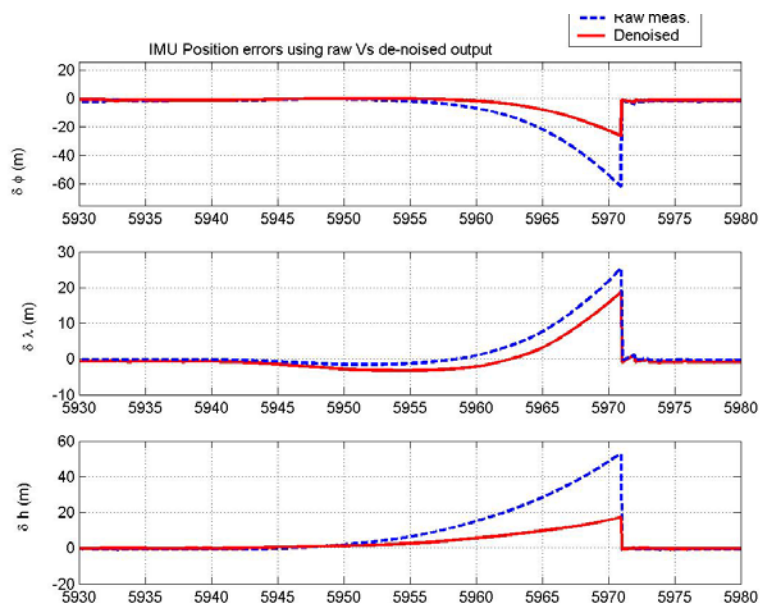


Figure 4.18 Position accuracy during GPS outage (3) – Raw versus de-noised INS measurements [denoising with soft threshold WMRA]

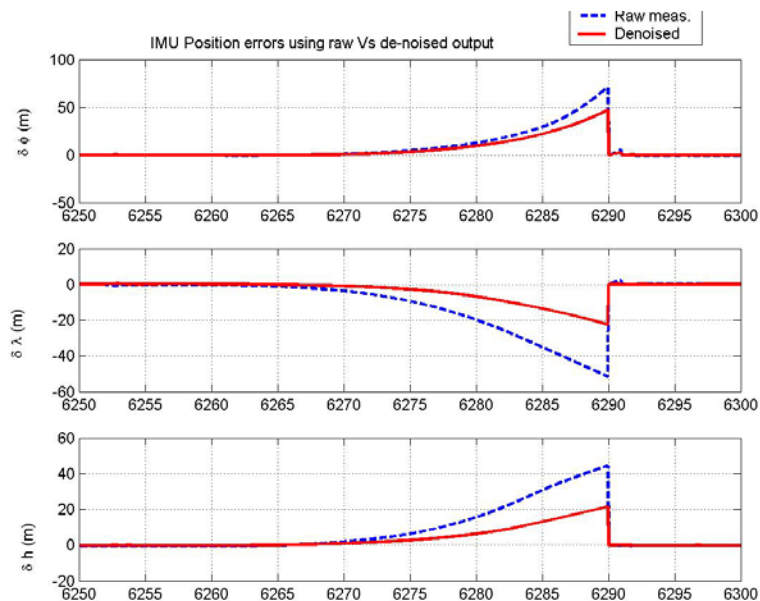


Figure 4.19 Position accuracy during GPS outage (4) – Raw versus de-noised INS measurements [denoising with soft threshold WMRA]

It can be clearly concluded from the above figures that significant improvements in the positioning accuracy can be achieved after applying the proposed WMRA method under the GPS outage scenarios described above. Positioning accuracy enhancement due to performing hard threshold WMRA is similar to that obtained by soft threshold WMRA implementation with a slight superiority for the latter thresholding technique. As discussed earlier in this section, this similarity situation is mainly due to the lack of sharp transitions in the real motion signal, which is traditionally the case in most urban land-vehicle navigations. Table 4.1 gives a numerical representation of the positioning accuracy enhancement after applying the proposed WMRA technique. The maximum error of each position component, typically achieved at the end of each outage period was

used as a measure of the performance of the KF before and after applying the proposed technique. Consequently, a rough estimate of positioning accuracy enhancement, as a percentage of the original error, can be obtained during the same above four graphically represented outages.

As a concluding remark, the above results and analysis have shown that the effectiveness of the proposed WMRA method under the ideal scenarios tested. It served as a pre-filtering stage before processing IMU data by the INS mechanization equations and the KF INS/GPS data fusion algorithm. The results of experimental validation, using real inertial data, indicated that significant improvement of about 50% in the positioning accuracy could be obtained after applying the proposed WMRA procedure. These initial promising results may pave the way for more future research work considering various conditions and motion scenarios in order to generalize such results. Should this happened, the MEMS-based systems could be involved in several navigation applications in the near future.

Table 4.1 Positioning accuracy enhancement after WMRA

	GPS outage (1)	GPS outage (2)	GPS outage (3)	GPS outage (4)	Average enhancement
Φ error before WMRA (m)	101.7	215.7	62.2	71.3	----
Φ error after WMRA (m)	45.2	13.3	26.6	47.2	----
Enhancement (%)	56 %	94 %	57 %	34 %	60 %
λ error before WMRA (m)	118.9	173.0	25.4	51.5	----
λ error after WMRA (m)	30.0	30.7	18.8	22.7	----
Enhancement (%)	75 %	82 %	26 %	56 %	59 %
Ht error before WMRA (m)	60.3	29.6	53.1	44.5	----
Ht error after WMRA (m)	55.0	23.9	17.2	21.5	----
Enhancement (%)	9 %	19 %	68 %	51 %	37 %

CHAPTER FIVE

FUZZY LOGIC BASED PREDICTION TECHNIQUES

As previously discussed in Chapter 3, the large bias and the highly random output drift of MEMS-based inertial sensors make it difficult to appropriately model either the deterministic or stochastic characteristics of these sensors through the traditionally used methods or techniques. This will significantly degrade the performance of the commonly used algorithms (e.g. Kalman filter) in INS and GPS integration particularly during the absence of a GPS solution. Therefore, other unconventional algorithms that are capable of modeling or predicting the effect of the high non-linearity of MEMS sensor output have to be implemented to overcome the above limitation. Networks-learning based algorithms, particularly those utilizing Fuzzy logic, are known for their efficiency in handling highly non-linear mathematical problems. In this Chapter, we present the fundamental basis of two proposed Fuzzy-based techniques to improve the integrated MEMS-INS/GPS system performance. The first proposed Fuzzy-based technique mainly adopts the Adaptive Neuro-Fuzzy Inference System (ANFIS) while the other utilizes the so-called Fuzzy Orthogonal Least-Squares (Fuzzy-OLS) algorithm. The two proposed Fuzzy techniques are mainly used to predict the MEMS-based IMU position and velocity errors during GPS signal outages where the conventional Kalman algorithm is working only in prediction mode.

5.1 Concepts of Fuzzy Logic

Information available for decision-making is usually not black and white; it generally involves some subtle “grey areas”. Fuzzy sets and fuzzy logic were developed to represent, manipulate, and utilize uncertain information and to develop a framework for handling uncertainty and imprecision in real-world applications [Lin et al., 2003]. Fuzzy logic systems provide an effective and accurate method for describing human perceptions. It accomplishes this by allowing computers to simulate human reasoning with less bias, and to behave with less analytical precision and logic than conventional computing methods [Turban and Aronson, 2000]. Moreover, Fuzzy logic with neural networks and probabilistic reasoning constitute the three cornerstones of soft computing, a trend that is growing in visibility [Zadeh, 1994]. Unlike the traditional hard computing techniques, such as using firm and precise mathematical formulas, soft computing strives to model the pervasive imprecision of the real world. Solutions derived from soft computing are generally more robust, flexible, and economical than those provided by hard computing [Malhotra and Malhotra, 1999].

Fuzzy logic provides an inference structure that enables approximate human reasoning capabilities to be applied to knowledge-based systems. The theory of fuzzy logic provides a mathematical strength to capture the uncertainties associated with human cognitive processes, such as thinking and reasoning. The conventional approaches to knowledge representation lack the means for representing the meaning of fuzzy concepts. As a consequence, the approaches based on first order logic and classical probability theory do not provide an appropriate conceptual framework for dealing with the

representation of commonsense knowledge, since such knowledge is by its nature both imprecise and non-categorical. The development of fuzzy logic was motivated by the need for a conceptual framework which can address the issue of uncertainty and imprecision. Some of the essential characteristics of fuzzy logic relate to the following [Zadeh, 1984]:

- Exact reasoning is viewed as a limiting case of approximate reasoning.
- Everything is a matter of degree.
- Knowledge is interpreted a collection of elastic or, equivalently, fuzzy constraint on a collection of variables.
- Inference is viewed as a process of propagation of elastic constraints.
- Any logical system can be fuzzified.

In addition, there are also two main characteristics of fuzzy systems that give them better performance for specific applications:

- Fuzzy systems are suitable for uncertain or approximate reasoning, especially for the system with a mathematical model that is difficult to derive.
- Fuzzy logic allows decision making with estimated values under incomplete or uncertain information.

In order to deal with imprecise or ill-defined systems, Fuzzy logic uses graded or quantified statements rather than ones that are strictly true or false. The fuzzy sets allow objects to have grades of membership of values ranging from zero to one. These sets, represented by linguistic variables, are used to represent a particular fuzzy set in a given

problem, such as “large”, ”medium” and “small” [Lin et al., 2003]. For instance, if U is a collection of objects denoted generically by $\{u\}$, which could be discrete or continuous, U is called the universe of discourse while u represents the generic element of U . A fuzzy set F in a universe of discourse U is characterized by a membership function μ_F which takes values in the interval $[0, 1]$. A fuzzy set F in U may be represented as a set of ordered pairs of a generic element u and its grade of membership function as follows:

$$F = \{(u, \mu_F(u)) \mid u \in U\}$$

$$F = \int_U \mu_F(u) / u \, du \quad (5.1)$$

$$F = \sum_{i=1}^n \mu_F(u_i) / u_i$$

Fuzzy sets have some operations that can be applied to crisp sets, as a subset of fuzzy sets. Assume that A and B are two fuzzy sets in U with membership function μ_A and μ_B , respectively. The set theory operation of union, intersection, complement and other relations of fuzzy sets are defined by their membership function. For example, the membership function $\mu_{A \cup B}$ of the union $A \cup B$ is point-wise defined for all $u \in U$ by:

$$\mu_{A \cup B}(u) = \max\{\mu_A(u), \mu_B(u)\} \quad (5.2)$$

While the membership function $\mu_{A \cap B}$ of the Intersection $A \cap B$ is point-wise defined for all $u \in U$ by:

$$\mu_{A \cap B}(u) = \min\{\mu_A(u), \mu_B(u)\} \quad (5.3)$$

The fuzzy inference system (FIS) is the process of formulating the mapping from a given input to an output using fuzzy logic. The dynamic behaviour of an FIS is characterized by a set of linguistic description rules based on expert knowledge. This expert knowledge is usually of the form [Tsoukalas and Uhrig, 1997]:

IF - a set of antecedent conditions is satisfied

THEN - a set of consequences can be inferred.

Since the antecedents and the consequent of these IF-THEN rules are associated with fuzzy concepts (linguistic terms), they are often called fuzzy conditional statements [Lee, 1990]. In the case of multiple-input-single-output (MISO) fuzzy system, the fuzzy rules have the form:

R_1 : if (x) is A_1 and (y) is B_1 then (z) is C_1 ,

R_2 : if (x) is A_2 and (y) is B_2 then (z) is C_2 ,

... ..

R_n : if (x) is A_n and (y) is B_n then (z) is C_n ,

where x, y, and z are linguistic variables representing two inputs process state variables and one output variable, respectively. A_i and B_i are linguistic values of the linguistic variables x, y in the universe of discourse U and V , respectively, with $i=1,2,\dots,n$. C_i are linguistic values of the linguistic variables z in the universe of discourse W in case of Mamdani FIS [Mamdani, 1974] and mapping linear functions of the inputs in case of Takagi-Sugeno (TS) FIS [Takagi and Sugeno, 1985].

5.2 Basic Elements of a Fuzzy Inference System (FIS)

The generic structure of an FIS consists of four main components as shown in Figure 5.1. The knowledge base (KB) stores the available knowledge about the problem in the form of Fuzzy “IF-THEN” rules. The other three components compose the Fuzzy inference engine that puts into effect the inference process on the system inputs by means of the Fuzzy rules. The fuzzification interface establishes a mapping between crisp values in the input domain (U) and Fuzzy sets defined in the same universe of discourse. The defuzzification interface realizes the opposite operation by establishing a mapping between Fuzzy sets defined in the output domain (V) and crisp values defined in the same universe. Further details and analysis on each of the above four components of FIS will be presented in the next sections [Babuska and Verbruggen, 1997].

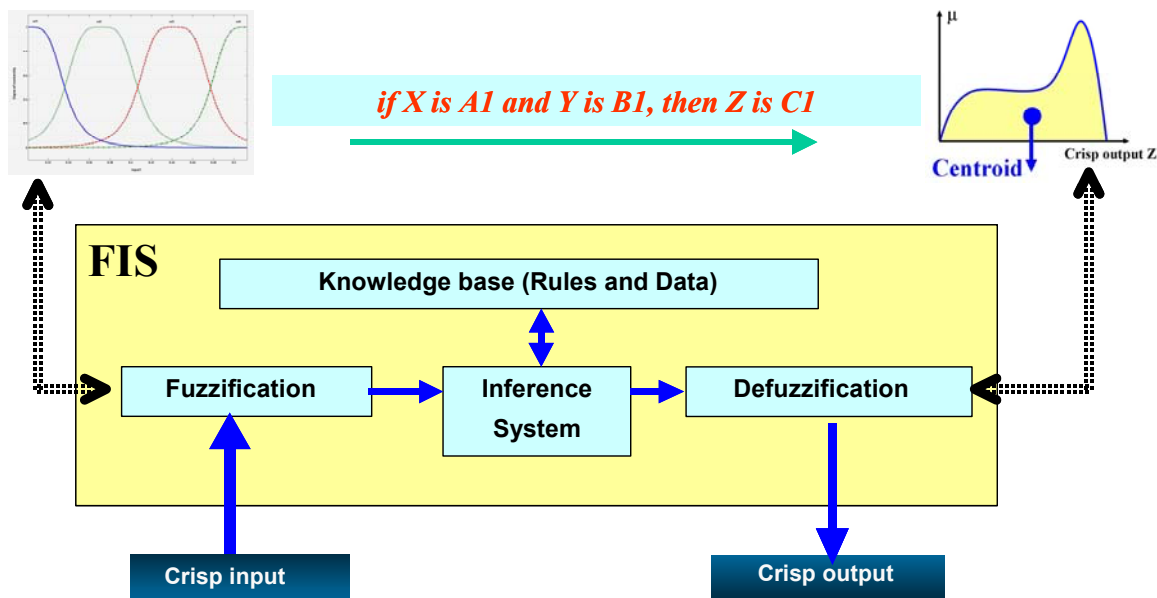


Figure 5.1 Basic components of a Fuzzy Inference System (FIS)

A. The knowledge base of Fuzzy rule-based systems

The KB establishes the fundamental functionality of the Fuzzy rule-based system. It serves as the repository of the problem specific knowledge, which models the relationship between input and output of the underlying system [Lee, 1990]. Upon this relationship, the inference process reasons from an observed input to an associated output. The KB contains two different information types, i.e., the Fuzzy rule semantics (in the form of fuzzy sets) and the linguistic rules representing the expert knowledge. This conceptual distinction is reflected by the two separate entities that constitute the KB [Mamdani, 1974]:

- A database (DB) containing the linguistic term sets considered in the linguistic rules and the membership functions defining the semantics of the linguistic labels. Each linguistic variable involved in the problem will have a Fuzzy partition of its domain associated to represent the Fuzzy set associated to each of its linguistic terms. Moreover, the DB also comprises the scaling factors or scaling functions that are used to transform between the universe of discourse in which the fuzzy sets are defined to the domain of the system input and output variables.
- A rule base (RB) is comprised by a collection of linguistic rules, usually in the form of “IF-THEN”, that are joined by the *also* operator. More explicitly, multiple rules can be fired simultaneously for the same input [Roubos et al., 2001].

B. The inference engine of Fuzzy rule-based systems

The inference engine of a fuzzy rule-based system is composed of the following three components [Tsoukalas and Uhrig, 1997]:

- A fuzzification interface that transforms the crisp inputs into Fuzzy values that serve as the input to the Fuzzy reasoning process.
- The inference system that infers from the Fuzzy input to several resulting output Fuzzy sets according to the information stored in the KB.
- A defuzzification interface that converts the Fuzzy sets obtained from the inference process into a crisp action that constitutes the global output of the Fuzzy rule-based system.

B.1. The fuzzification inference

The fuzzification inference enables the Fuzzy rule-based system to handle crisp input values. Fuzzification establishes a mapping from crisp input values to Fuzzy sets defined in the universe of discourse of that input. In other words, it performs a scale mapping that transfers the range of values of input variables into corresponding universes of discourse. The main function of fuzzification is to convert input data into suitable linguistic values which may be viewed as labels of fuzzy sets. The most commonly used functions for fuzzification are the Bell-shaped and the Gaussian distribution [Mamdani, 1974].

B.2. The inference system

The inference system is the component that derives the Fuzzy outputs from the input Fuzzy sets according to the relation defined through Fuzzy rules [Tsoukalas and Uhrig,

1997]. The inference scheme establishes a mapping between Fuzzy sets $U = U_1, U_2, \dots, U_n$ in the input domain of $X = X_1, \dots, X_n$ and fuzzy sets V in the output domain of Y . The concepts and algorithms used by an FIS will be explained and analysed in more details later in the following section.

B.3. The defuzzification inference

The inference process in the Fuzzy rule-based system operates on the level of individual rules. Therefore, the application of the computational rule of inference to the current input using m number of rules in the KB generates m output Fuzzy sets. The defuzzification interface has to aggregate the information provided by the m output Fuzzy sets and to obtain a crisp output value from them. The contribution of each Fuzzy set is considered separately and the final crisp value is obtained by means of an averaging or selection operation performed on the set of crisp values derived from each of the individual Fuzzy sets. The most common choice is the “centre of area or gravity” (CG) or the “maximum value” (MV) weighted by the matching degree using the following expression [Mamdani and Assilian, 1974]:

$$y_o = \frac{\sum_{i=1}^m h_i y_i}{\sum_{i=1}^m h_i} \quad (5.4)$$

where y_i is the CG or the MV of the Fuzzy set inferred from rule # i and h_i represents the degree of match between the actual system input and the rule antecedent.

As a concluding remark for this section, it is worthwhile to mention that in case of implementing a Takagi-Sugeno (TS) Fuzzy rule-based system, the consequent part of the rule should be represented as a function of the input variables. The most common form of these kinds of rules is the one in which the consequent expression constitutes a linear combination of the variables involved in the antecedent part of the rule [Sugeno, 1985]. The output of the TS Fuzzy rule-based system using a KB composed of m rules is obtained as a weighted sum of the individual outputs provided by each rule (y_i), $i=1, \dots, m$. In this case the terms of Equation 5.4 are as follows:

$h_i = T(A_{i,1}(x_1), \dots, A_{i,n}(x_n))$ is the matching degree between the antecedent part of the rule # i and the current inputs to the system ($x_o = x_1, \dots, x_n$). T stands for a conjunctive operator modelled by a t-norm with the most common choices being the minimum or the algebraic product.

5.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

In order for an FIS to be mature and well established so that it can work appropriately in prediction mode, its initial structure and parameters (linear and non-linear) need to be tuned or adapted through a learning process using a sufficient input-output pattern of data. One of the most commonly used learning systems for adapting the linear and non-linear parameters of an FIS, particularly TS type, is the ANFIS. ANFIS is a class of adaptive networks that are functionally equivalent to fuzzy inference systems [Jang, 1993]. The architecture of ANFIS and the main concepts and algorithms adopted during its learning process are described hereafter.

For simplicity, an FIS structure of two inputs x and y and a single output z is considered. Assume that the rule base contains only two Fuzzy if-then rules of TS type [Takagi and Sugeno, 1985] as follows:

Rule (1): If X is A_1 and Y is B_1 , then $Z_1 = m_1 X + n_1 Y + q_1$,

Rule (2): If X is A_2 and Y is B_2 , then $Z_2 = m_2 X + n_2 Y + q_2$.

The architecture of the proposed FIS (see Figure 5.2) contains five layers where the node functions in the same layer are of the same function family. Inputs, outputs and implemented mathematical models of the nodes of each layer are explained below.

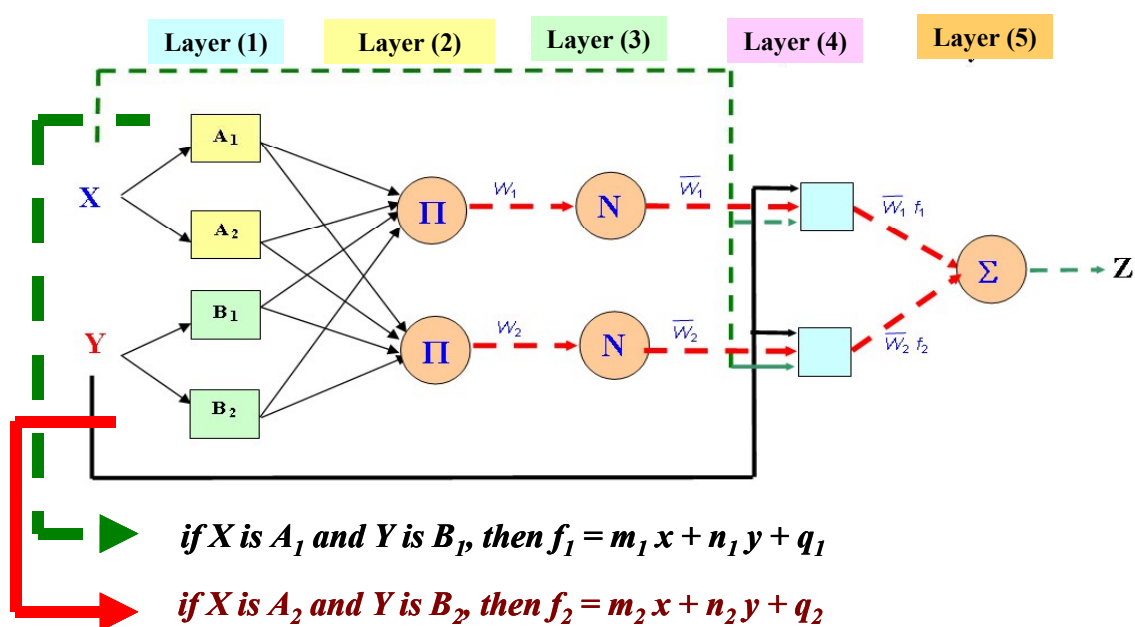


Figure 5.2. A simple TS-FIS structure of two inputs, two rules and a single output [after Reda-Taha et al., 2003]

Layer 1 The node function of every node i in this layer take the form:

$$O_i^1 = \mu A_i (X) \quad (5.5)$$

where X is the input to node i , μA_i is the membership function of the linguistic label A_i associated with this node and O_i is the degree of match to which the input X satisfies the quantifier A_i . Usually $\mu A_i (X)$ is chosen to be a bell-shaped (Equation 5.6) or Gaussian function (Equation 5.7) with maximum equal to 1 and minimum equal to 0, such as:

$$\mu A_i (X) = \frac{1}{1 + \left[\left(\frac{X - c_i}{a_i} \right)^2 \right]^{b_i}} \quad (5.6)$$

$$\mu A_i (X) = \exp \left\{ -\frac{1}{2} \frac{(X - v_i)^2}{\sigma_i^2} \right\} \quad (5.7)$$

where $\{a_i, b_i, c_i, v_i, \sigma_i^2\}$ is the premise non-linear parameter set. As the value of these parameters change, either the bell-shaped or Gaussian function varies accordingly thus exhibiting various forms of membership functions on the linguistic label A_i . Moreover, some other continuous and piecewise differentiable functions, such as trapezoidal and triangular membership functions are also qualified candidates for node functions in this layer.

Layer 2 Every node in this layer multiplies the incoming signals from layer 1 and sends the product out as follows,

$$w_i = \mu A_i (X) \times \mu B_i (Y), \quad i = 1, 2 \quad (5.8)$$

where the output w_i represents the firing strength of a rule. Beside multiplication, other T-norm operators that perform generalized and can be also implemented as the node function in this layer.

Layer 3 Every node i in this layer determine the ratio of the i -th rule's firing strength to the sum of all rules' firing strengths as:

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2 \quad (5.9)$$

where the output of this layer represent the normalized firing strengths.

Layer 4 Every node i in this layer has a node function of the form:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (m_i X + n_i Y + q_i) \quad (5.10)$$

where \bar{w}_i is the output to layer 3, and $\{m_i, n_i, q_i\}$ is the linear consequent parameter set.

Layer 5 There is only a single node in this layer that computes the overall output as the weighted average of all incoming signals from layer 4 as:

$$O_i^5 = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (5.10)$$

As mentioned earlier, both the premise (non-linear) and consequent (linear) parameters of the FIS should be tuned, utilizing the so-called learning process, to optimally represent the factual mathematical relationship between the input space and output space. Normally, as a first step, an approximate fuzzy model is initiated by the system and then

improved through an iterative adaptive learning process. The training algorithm, namely ANFIS, was developed by Jang [1993]. Basically, ANFIS takes the initial fuzzy model and tunes it by means of a hybrid technique combining gradient descent back-propagation and mean least-squares optimization algorithms (see Figure 5.3). At each epoch, an error measure, usually defined as the sum of the squared difference between actual and desired output, is reduced. Training stops when either the predefined epoch number or error rate is obtained. The gradient descent algorithm is mainly implemented to tune the non-linear premise parameters while the basic function of the mean least-squares is to optimize or adjust the linear consequent parameters.

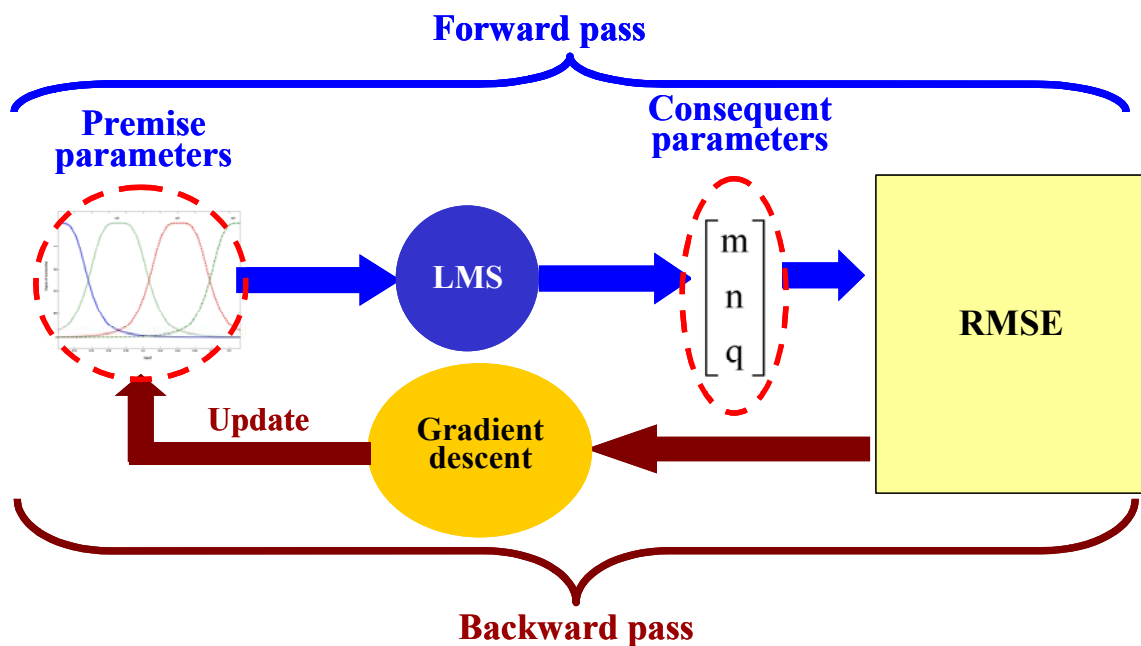


Figure 5.3 ANFIS learning using hybrid technique

5.3.1 Consequent Parameters Estimation

According to the proposed simple ANFIS architecture (shown in Figure 5.2), it can be observed that given the values of the premise parameters, the overall output can be expressed as a linear combination of the consequent parameters. More explicitly, the output Z in Figure 5.2 can be rewritten as:

$$\begin{aligned} Z &= \bar{w}_1 f_1 + \bar{w}_2 f_2 \\ &= (\bar{w}_1 X) m_1 + (\bar{w}_1 Y) n_1 + (\bar{w}_1) q_1 + (\bar{w}_2 X) m_2 + (\bar{w}_2 Y) n_2 + (\bar{w}_2) q_2 \end{aligned} \quad (5.11)$$

By fixing the elements included in the normalized firing strength of the rules \bar{w}_1 and \bar{w}_2 , the above expression in Equation 5.11 is linear in the consequent parameters ($m_1, n_1, q_1, m_2, n_2, \text{ and } q_2$). Therefore, the matrix form of Equation 5.11 can be written as:

$$A X = B \quad (5.12)$$

where X is an unknown vector whose elements are the linear consequent parameters of all rules. Given that the number of these linear consequent parameters is (M) and the number of training input output data pairs is (N), then the dimensions of X , A and B are $N \times M$, $M \times I$ and $N \times I$, respectively. Since N is always greater than M , the system of equations in (5.12) is over-determined and generally there is no exact or unique solution that can be attained. Instead, a least-squares estimate (LSE) of X is sought to minimize the squared error $\|AX - B\|^2$. This is a standard problem that forms the base in many applications like linear regression, adaptive filtering and signal processing. The most well known formula

for solving such over-determined system of equations uses the pseudo-inverse of X [Astrom and Wittenmark, 1984]:

$$X^* = (A^T A)^{-1} A^T B \quad (5.13)$$

where A^T is the transpose of A and $(A^T A)^{-1} A^T$ is the pseudo inverse of A if $(A^T A)$ is non-singular. While Equation 5.13 is concise in notation, it is expensive in computation when dealing with the matrix inverse and, moreover, it becomes ill-defined if $(A^T A)$ is singular. In order to avoid the high computation load or the singularity problem, sequential formulas are employed to compute the least-squares estimate of X . This sequential method is more efficient, particularly when M is small, and can be easily modified to an on-line version for systems with changing characteristics. Specifically, if the i -th row vector of matrix A in Equation 5.12 is denoted by a_i^T and the i -th element of the vector B is denoted by b_i^T , then X can be calculated iteratively using the following sequential formulas [Astrom and Wittenmark, 1984] [Fahlman, 1988]:

$$X_{i+1} = X_i + S_{i+1} a_{i+1} (b_{i+1}^T - a_{i+1}^T X_i) \quad (5.14)$$

$$S_{i+1} = S_i - \frac{S_i a_{i+1} a_{i+1}^T S_i}{1 + a_{i+1}^T S_i a_{i+1}}, \quad i = 0, 1, \dots, N-1$$

where S_i is often called the covariance matrix and the least squares estimate X^* is equal to X_N . The initial condition to bootstrap Equation 5.14 are $X_0 = 0$ and $S_0 = \eta I$, where η is an arbitrary positive large number and I is the identity matrix of dimension $M \times M$.

5.3.2 Antecedent Parameters Tuning

After performing the LSE for the optimal estimation of the FIS consequent parameters, the residuals are estimated and therefore the root-mean-square-error (RMSE) can be determined. This RMSE value is used in a gradient-descent based back propagation approach [Jang, 1993] in order to tune the premise or antecedent non-linear parameters as will be described hereafter. Generally, if an adaptive network has L layers and the K -th layer has $\#(k)$ nodes, the i -th position of the k -th layer can be denoted by (k, i) and its node function (or node output) by O_i^k . Since a node output depends on its incoming signals and its parameters set, it can be expressed as:

$$O_i^k = O_i^k (O_i^{k-1}, \dots, O_{\#(k-1)}^{k-1}, a, b, c, \dots) \quad (5.15)$$

where a, b, c , etc. are the parameters pertaining to this node.

Assuming the given training data set has P input-output entries, the error measure for the p -th entry ($1 \leq p \leq P$) of training data entry can be defined as the sum of squared errors [Jang, 1991]:

$$E_p = \sum_{m=1}^{\#(L)} (T_{m,p} - O_{m,p}^L)^2 \quad (5.16)$$

where $T_{m,p}$ is the m -th component of the p -th target output vector, $\#(L)$ is the number of outputs at the last layer L , and $O_{m,p}^L$ is the m -th component of actual output vector produced by the presentation of the p -th input vector. Hence the overall measure is:

$$E = \sum_{p=1}^P E_p .$$

In order to develop a learning procedure that implements gradient descent in E over the parameters' space, first the error rate $\frac{\partial E_p}{\partial O}$ has to be calculated for the p -th training data and for each node output O . The error rate for the output node at (L, i) can be calculated readily from Equation 5.16:

$$\frac{\partial E_p}{\partial O_{i,p}^L} = -2 (T_{i,p} - O_{i,p}^L) \quad (5.17)$$

For the internal node at (k, i) , the error rate can be derived by the chain rule [Werbos, 1974] as:

$$\frac{\partial E_p}{\partial O_{i,p}^L} = \sum_{m=1}^{\#(k+1)} \frac{\partial E_p}{\partial O_{m,p}^{k+1}} \frac{\partial O_{m,p}^{k+1}}{\partial O_{i,p}^k} \quad (5.18)$$

where $1 \leq k \leq L-1$. That is the error rate of an internal node can be expressed as a linear combination of the error rates of the nodes in the next layer. Therefore for all $1 \leq k \leq L$ and $1 \leq i \leq \#(k)$, $\frac{\partial E_p}{\partial O_{i,p}^k}$ can be found using Equations 5.17 and 5.18.

Now, if a is a parameter of the given adaptive network, we have:

$$\frac{\partial E_p}{\partial a} = \sum_{o^* \in S} \frac{\partial E_p}{\partial O^*} \frac{\partial O^*}{\partial a} \quad (5.19)$$

where S is the set of nodes whose outputs O depend on a . Then the derivative of overall error measure E with respect to a is:

$$\frac{\partial E}{\partial a} = \sum_{p=1}^P \frac{\partial E_p}{\partial a} \quad (5.20)$$

Accordingly, the update formula for the generic parameter a is:

$$\Delta a = -\varepsilon \frac{\partial E}{\partial a} \quad (5.21)$$

in which ε is a learning rate which can be further expressed as:

$$\varepsilon = \frac{\kappa}{\sqrt{\sum_a \left(\frac{\partial E}{\partial a}\right)^2}} \quad (5.22)$$

where κ is known as the training step size which defines the length of each gradient transition in the parameter space. Usually, the value of κ is changed to vary the speed of convergence of the FIS model [Werbos, 1974].

ANFIS, as a class of adaptive networks, can apply all the above illustrated gradient descent concepts and algorithms (Equation 5.15 to Equation 5.22). However, two particular characteristics for ANFIS should be considered. First, the total number of layers L is explicitly fixed at five (see Figure 5.2) with the number of nodes in each layer depending mainly on the number of inputs, the number of clusters associated with each input and the number of fuzzy rules. Second, there is only a single output at the fifth and last layer, i.e. $\#(L) = 1$. Fortunately, these two particular characteristics of ANFIS will even add more simplicity when implementing the previously described gradient-descent method to tune the antecedent parameters.

5.3.3 Limitation of ANFIS

The required time for a successful ANFIS learning, in order to achieve FIS model convergence, is always a matter of concern. Such concern becomes more critical when performing ANFIS training adaptively to model the behaviour of highly non-linear dynamic systems (e.g. low cost or MEMS-based INS). In this case a sliding window, comprising a short-term sequence of input-output data pairs, is used for continuously tuning the FIS parameters in an adaptive manner. The short length of the sliding window will help decrease the number of data entries and thus decreasing the required time for system convergence. Nevertheless, a large number of Fuzzy if-then rules, used to represent the system dynamics, will significantly increase the convergence time such that the ANFIS performance might not be reliable in real-time applications. Mainly, the number of rules of the FIS, initiated by the system, depends on the overall number of input Fuzzy clusters [Abonyi et al., 2001]. More explicitly, each possible combination of a subset of Fuzzy clusters will result in a new rule. For instance, if the FIS has 10 inputs and the fuzzified space of each of them is classified into five Fuzzy sets or clusters, then the resultant possible number of rules that will be initiated by the system, is $(5^{10} = 9765625)$. This large number of rules will disrupt the performance of ANFIS training such that it will be lag behind the requirements of real-time applications. In order to overcome this problem, an efficient algorithm has to be first implemented offline using a sufficient amount of input-output data pairs for the effective compaction of the initial FIS size. In this regard, an offline technique, namely Fuzzy-Orthogonal-Least-Squares (Fuzzy-OLS), is proposed in this thesis. The main concepts and mathematical foundation

of Fuzzy-OLS technique will be illustrated in details in the remaining part of this Chapter.

5.4 Principles of Fuzzy OLS

Different approaches to obtain TS-fuzzy models from data have been proposed. Most approaches, however, utilize only the function approximation capabilities of fuzzy systems, and little attention is paid to the qualitative aspects. This makes them less suited for applications in which emphasis is not only on accuracy, but also on interpretability, computational complexity and training time [Setnes et al., 1998]. Additionally, such straightforward approaches soon lead to quite complex models because generally little effort is put into variable selection and in the creation of suitable local rules. Moreover, in general no model reduction is applied, while this may simplify the model by removing redundant information. Abonyi et al. [2001] have proposed a combined Fuzzy-OLS method that handles these issues in order to create compact TS models that can be effectively used to represent complex systems.

The bottleneck of the TS model identification is the data-driven identification of the structure that requires non-linear optimization. For this purpose, usually heuristic (data-driven) approaches, like fuzzy clustering methods, are applied mainly for determining the rule antecedents of TS models [Babuska, 1998]. Fuzzy-OLS method utilizes a more advanced clustering method that goes a step further in accomplishing the total parameter and structure identification of TS models. This approach is based on a new clustering algorithm obtained by modifying the Gath-Geva clustering [Abouni et al., 2001a]. The

clusters are formed by expectation maximization (EM) of TS model identification in the form of a mixture of Gaussian models [Bishop, 1995]. The obtained model is then reduced by reducing the amount of antecedent variables and also the amount of consequent variables. Using too many antecedent variables may result in difficulties in the prediction and interpretability capabilities of the model due to redundancy, non-informative features and noise. Hence, the proper selection of these variables is usually necessary. For this purpose, a modified version of the Fischer Interclass Separability method can be employed for the selection of the most effective fuzzy classifiers [Roubos et al., 2001].

An Orthogonal Least Squares (OLS) method is proposed for reduction of the consequent space. The application of orthogonal transforms for the reduction of the number of rules has received much attention in recent literature [Yen and Wang, 1999; Setnes and Hellendoorn, 2000]. These methods evaluate the output contribution of the rules to obtain an importance ordering. As will be discussed later, OLS can be applied for the selection of the most relevant input and consequent variables based on the OLS analysis of the local models of the clusters. An effective clustering algorithm that allows for the direct identification of TS models is illustrated first in Section 5.4.1. Then, the selection of consequent variables by OLS based method as well as the selection of the antecedent variables, based on Fisher's Interclass Separability criteria is discussed in Section 5.4.2.

5.4.1 TS Model Identification Through Optimal Clustering

The main objective of ANFIS learning process is to develop an algorithm for the identification of unknown non-linear systems of the form:

$$y_k = f(\mathbf{u}_k) \quad (5.23)$$

based on a specified or measured input data $\mathbf{u}_k = [u_{1,k}, \dots, u_{m,k}]^T$ and measured output data y_k of the system where $k = 1, \dots, N$ denotes the index of the k -th input output data pair [Murray-Smith and Johansen, 1997]. In addition, the operating regime of the TS Fuzzy model can be formulated as:

$$\hat{y}_k = \sum_{i=1}^r \bar{w}_i(x_k) (m_i^T \phi_k + q_i) \quad (5.24)$$

where $\bar{w}_i(x_k)$ describes the operating regime of the i -th local linear model defined by the parameter vector $\theta_i = [m_i^T \ q_i]^T$, $i = 1, \dots, r$, ϕ_k is the regression vector, where x_k and ϕ_k are n and n_r dimensional subsets of the original input vector \mathbf{u}_k , respectively [Takagi and Sugeno, 1985]. Note that Equation 5.24 represents the generalized form of Equation 5.10.

The objective of clustering is to partition the identification data Z into c clusters. This means, each observation consists of the input and the output variables, grouped into a row vector $z_k = [\mathbf{u}_k, y_k]$, where k denotes the k -th row of the Z matrix. The fuzzy partition is represented by the $U = [\mu_{i,k}]_{c \times N}$ matrix, where the $\mu_{i,k}$ element of the matrix represents the degree of membership, how the z_k observation is in the cluster $i = 1, \dots, c$.

The suggested fuzzy clustering method utilizing an adapted version from the Gath-Geva method [Abonyi et al., 2001a] is now described. Each cluster contains an input distribution, a local model and an output distribution to represent the density of the data. The input distribution is parameterized as an unconditional Gaussian [Gershenfeld et al., 1999] distribution, and defines the domain of influence of a cluster similarly to multivariate membership functions. The clustering is based on the minimization of the sum of weighted squared distances between the data points z_k and the cluster prototypes η_i that contain the parameters of the clusters. The square of the $D_{i,k}^2$ distances are weighted with the membership values $\mu_{i,k}$ in the objective function that is minimized by the clustering algorithm and formulated as [Turčan et al., 2003]:

$$J(\mathbf{Z}, \mathbf{U}, \boldsymbol{\eta}) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{i,k}) D_{i,k}^2 = \sum_{i=1}^c \sum_{k=1}^N (\mu_{i,k}) \|\mathbf{z}_k - \mathbf{v}_i\|^2 \quad (5.25)$$

where $D_{i,k}^2 = \|\mathbf{z}_k - \mathbf{v}_i\|^2$ is dissimilarity measure between the sample z_k and the centre v_i of the specific cluster i (Euclidean distance). To get a Fuzzy partitioning space, the membership values have to satisfy the following conditions:

$$U \in R^{c \times N} \left| \mu_{i,k} \in [0,1], \forall i, k; \quad \sum_{i=1}^c \mu_{i,k} = 1 \forall k; \quad 0 \leq \sum_{k=1}^N \mu_{i,k} \leq N \forall i \right. \quad (5.26)$$

The minimization of Equation 5.25 function represents the non-linear optimization problem that is subject to constraints defined by Equation 5.26 and can be solved by using a variety of available methods. The most popular method is the alternating optimization (AO), which is formulated as shown below.

Given a set of data Z , the number of clusters c is specified as well as the termination tolerance $\varepsilon > 0$. The partitioning matrix $U = [\mu_{i,k}]_{c \times N}$ is initialized randomly where $\mu_{i,k}$ denotes the membership or the degree of match of z_k input-output data to the i -th Fuzzy cluster. First, the parameters of the clusters are calculated including centres and standard deviation of the Gaussian membership functions:

$$\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^N \mu_{i,k}^{(l-1)} \mathbf{x}_k}{\sum_{k=1}^N \mu_{i,k}^{(l-1)}}, \quad \sigma_{i,j}^{2(l)} = \frac{\sum_{k=1}^N \mu_{i,k}^{(l-1)} (x_{j,k} - v_{i,j})^2}{\sum_{k=1}^N \mu_{i,k}^{(l-1)}}, \quad 1 \leq i \leq c \quad (5.27)$$

Then, the parameters of local models are estimated by weighted least squares approach. This approach doesn't estimate all parameters simultaneously. Instead, the parameters of local models are estimated separately using a set of local estimation criteria [Babuska, 1998]:

$$\min_{\theta_i} \|y - \phi_e \theta_i\|^2 \beta_i = \left\| y \sqrt{\beta_i} - \phi_e \sqrt{\beta_i} \theta_i \right\|^2 \quad (5.28)$$

where $\|y - \phi_e \theta_i\|^2$ represents the difference between the actual and the estimated fuzzy output. ϕ_e denotes the extended regression matrix obtained by adding a unitary column to ϕ , $\phi_e = [\phi \quad 1]$, where $\phi = [\phi_1, \dots, \phi_N]^T$, $y = [y_1, \dots, y_N]^T$, and β_i denotes a diagonal matrix having membership degrees in its diagonal elements:

$$\beta_i = \begin{bmatrix} \mu_{i,1} & 0 & \dots & \dots & 0 \\ 0 & \mu_{i,2} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & \mu_{i,N} \end{bmatrix} \quad (5.29)$$

The weighted least-squares estimate of the rule consequent parameters is given by:

$$\theta_i = (\phi_e^T \beta_i \phi_e)^{-1} \phi_e^T \beta_i y \quad (5.30)$$

The standard deviation of the modeling error can be expressed as:

$$\sigma_i^2 = \frac{\sum_{k=1}^N (y_k - f_i(\phi_k, \theta_i))^T (y_k - f_i(\phi_k, \theta_i)) \mu_{i,k}^{(l-1)}}{\sum_{k=1}^N \mu_{i,k}^{(l-1)}} \quad (5.31)$$

The following step is to calculate the distance measure $D_{i,k}^2(z_k, \eta_i)$. This measure consists of two terms; the first term is based on the geometrical distance between the v_i cluster centre and x_k , while the second is based on the performance of the local linear models [Babuska, 1998]:

$$\frac{1}{D_{i,k}^2(z_k, \eta_i)} = \prod_{j=1}^n \exp\left(-\frac{1}{2} \frac{(x_{j,k} - v_{i,j})^2}{\sigma_{i,j}^2}\right) \frac{\exp\left(-\frac{(y - [\phi \ 1]\theta_i^T)^T (y - [\phi \ 1]\theta_i^T)^T}{\sqrt{2\pi\sigma_i^2}}\right)}{\sqrt{2\pi\sigma_i^2}} \quad (5.32)$$

The above distance measure is used to update the partition matrix U :

$$\mu_{i,k}^l = \frac{1}{\sum_{i=1}^c (D_{i,k}^2(z_k, \eta_i) / D_{i,k}^2(z_k, \eta_i))^2}, \quad 1 \leq k \leq N \quad (5.33)$$

This sequence of operations is iterated till the termination tolerance is achieved:

$$\|U^{(l)} - U^{(l-1)}\| < \varepsilon \quad (5.34)$$

5.4.2 Consequent and Antecedent Parameters Selection

As discussed, having too many antecedent and consequent variables results in difficulties in the prediction and interpretability capabilities of the fuzzy model due to redundancy, non-informative features and noise. To avoid these problems, two methods are presented in this section; the first for selecting the most effective consequent variables while the second for filtering the antecedent variables based on their importance. Such reduction of the number of antecedent and consequent parameters is very useful and results in a compacted FIS structure.

I. Selection of the Consequent Variables by Orthogonal Least Squares (OLS) Method

As the fuzzy model is linear in the parameters θ_i , the solution for Equation 5.28 can be formulated as:

$$\theta_i = B^+ y \sqrt{\beta_i} \quad (5.35)$$

where B^+ denotes the Moore-Penrose pseudo inverse of $\phi_e \sqrt{\beta_i}$ [Billings et al., 1988].

The OLS method transforms the columns of B into a set of orthogonal basis vectors in

order to inspect the individual contribution of each rule. To do this, Gram-Schmidt orthogonalization of $(B=WA)$ is used, where W is an orthogonal matrix ($W^T W = I$) and A is the upper triangular matrix with unity diagonal elements [Yen and Wang, 1999]. If w_i denotes the i -th column of W and g_i is the corresponding element of the OLS solution vector $g = A\theta_i$, the output variance $(y\sqrt{\beta_i})^T (y\sqrt{\beta_i}) / N$ can be explained by the regressors $(y\sqrt{\beta_i})^T (y\sqrt{\beta_i}) / N$. Thus the error reduction ratio, ρ , due to an individual rule i can be expressed as [Setnes and Hellendoorn, 2000]:

$$\rho^i = \frac{g_i^2 w_i^T w_i}{(y\sqrt{\beta_i})^T (y\sqrt{\beta_i})} \quad (5.36)$$

This error reduction ratio offers a simple mean of ordering the rules and consequently the corresponding consequent variables. Therefore, it can be easily used to select only the most influencing subset of input in a forward regression manner.

II. Selection of The Antecedent Variables by Interclass Separability

Antecedent variables selection is usually necessary. Abonyi et al [2001] have modified the Fischer Interclass Separability method, which is based on statistical properties of the data. The modified version has been applied for feature selection of premise Fuzzy clusters [Roubos et al., 2001]. The interclass separability criterion is based on the F_B between-class and the F_W within-class covariance matrices that sum up to the total covariance of the training data F_T , where:

$$F_W = \sum_{i=1}^c p(\eta_i) F_i, \quad F_B = \sum_{i=1}^c p(\eta_i) (v_i - v_o)^T (v_i - v_o) \quad (5.37)$$

where $p(\eta_i) = \frac{1}{N} \sum_{k=1}^N \mu_{i,k}$, $v_o = \sum_{i=1}^c p(\eta_i)$, and F_i is a diagonal matrix that contains the variances of the Gaussian membership functions.

The feature interclass separability selection criterion is a trade-off between the F_W and F_B :

$$J = \frac{\det(F_B)}{\det(F_W)} \quad (5.38)$$

The importance of a feature (antecedent Fuzzy cluster) is measured by leaving out the feature and calculating J for the reduced covariance matrices. The feature selection is made iteratively by leaving out the least needed feature [Jang, 1996].

5.5 Summary

The Fuzzy-OLS and ANFIS techniques were analysed conceptually and mathematically, in details, since they will be utilized later on in the two subsequent Chapters of this thesis. First, the ANFIS is proposed to model the non-linear characteristics of MEMS-based inertial sensors particularly the thermal dependant variation of either accelerometer or gyro output (Chapter 6). Second, a combined Fuzzy-OLS and ANFIS-based network is suggested to be augmented with the conventional KF solution of MEMS-INS/GPS integrated system (Chapter 7). The reason beyond the latter suggested augmentation is to provide more appropriate and reliable position and velocity estimates during GPS outage, i.e. in standalone inertial navigation mode. Normally, an offline Fuzzy-OLS training is

implemented in order to develop an initial a simple and compact TS FIS structure that can be adaptively tuned on-line through ANFIS and a sliding window of input-output data pairs.

CHAPTER SIX

THERMAL CHARACTERISTICS OF MEMS INERTIAL SENSORS

The miniature size and fabrication process of MEMS-based inertial sensors make them highly sensitive to the surrounding environment. MEMS-based sensors' performance characteristics (e.g. bias and scale factor) vary significantly with temperature changes which introduce additional errors to their output. Such errors, unless modeled and compensated, will lead to a significant degradation in the accuracy of final estimated navigation states (position, velocity and attitude). In this Chapter, the effect of thermal variation on MEMS-based inertial sensors is first discussed. The results of lab thermal testing of two MEMS-based IMUs (Crossbow AHRS400Ca and MotionPak II) are presented. Finally, a proposed ANFIS model of the MEMS inertial output variation is illustrated and experimentally validated.

6.1 Thermal Characteristics and Variation of MEMS Inertial Sensors Output

There are many advantages to having miniature sensors and systems. From a practical aspect, smaller often implies easier to carry and low-power, features that are crucial in the case of a personal and land vehicle navigation systems. From a physical aspect, smaller and lighter can also mean higher resonant frequencies (and in turn higher operating frequencies and bandwidths) and shorter thermal time constants. With recent advances in microelectronics, it is possible to combine intelligent electronics on the same sensor chip, resulting in relatively complex and advanced systems. Examples of this are closed loop

accelerometers and complete single chip IMU designs [Schwarz and El-Sheimy, 1999]. On the negative side, the miniature size and lightweight of MEMS inertial sensors can increase their sensitivity to temperature variation. A smaller proof mass also introduces additional noise, known as thermal noise, on the raw output of MEMS sensors [Lingjun et al., 2001]. As a result, the output measurements of these low cost sensors may drift significantly with the variation of the surrounding environmental temperature [HuiPing et al., 2001]. This will effectively diminish the resulting positioning accuracy through integrating the thermally defective linear acceleration and angular rate measurements. Figure 6.1 shows the thermal drift of the MEMS Cross-bow AHRS300CA Z-axis accelerometer's output as an example. The MEMS Cross-bow unit was mounted inside a climatic chamber where the temperature was kept fixed for half an hour at different temperature points, from $-30\text{ }^{\circ}\text{C}$ to $+30\text{ }^{\circ}\text{C}$ with $5\text{ }^{\circ}\text{C}$ increments. The output of the Z-axis accelerometer was recorded at each temperature point during the 30 minutes of temperature stability. Note that the raw output sequence is presented in the upper plot of the figure while the de-noised output using WMRA (see Chapter 4) is illustrated in the lower plot. The smoothness or noise reduction gained through WMRA implementation is clearly noticed through the significant decrease in the variance values. Figure 6.2 illustrates the same test results for the Z-axis gyroscope's output instead.

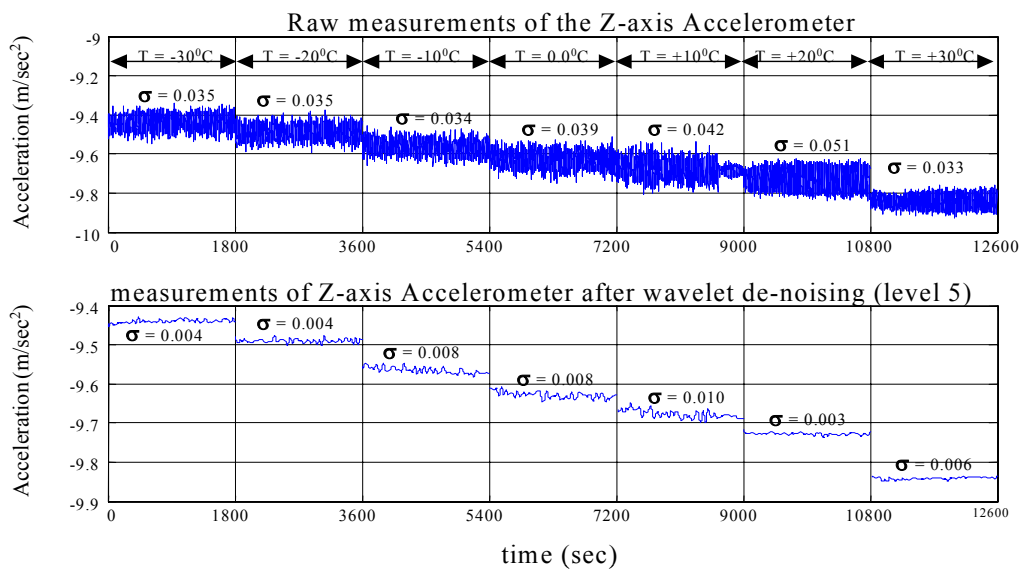


Figure 6.1 Extracting the thermal trend of the Cross-Bow AHRS300CA Z-axis accelerometer using wavelet decomposition

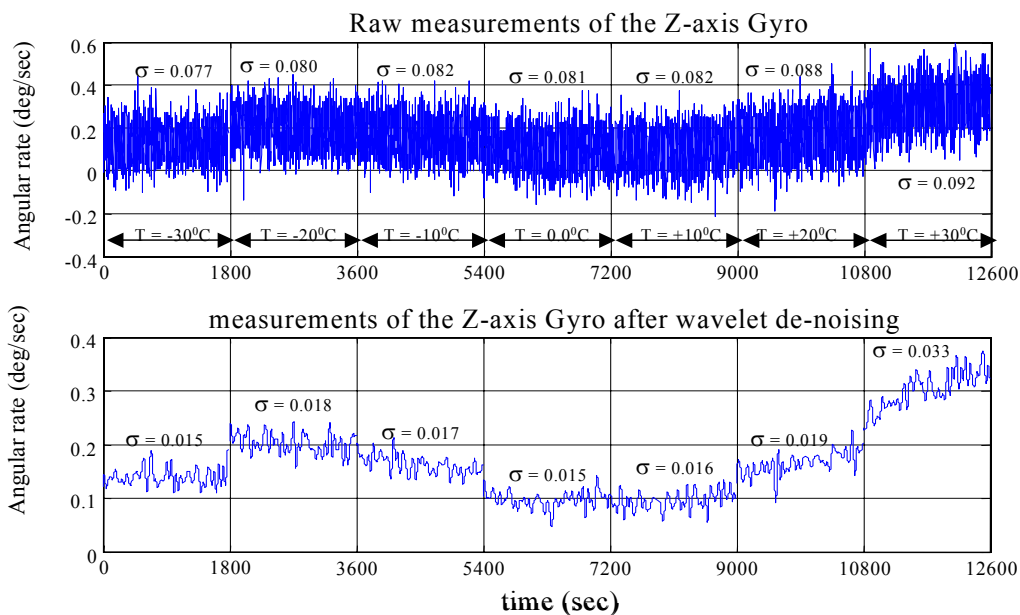


Figure 6.2 Extracting the thermal trend of the Cross-Bow AHRS300CA Z-axis gyroscope using wavelet decomposition

Normally, the effect of temperature change on the MEMS-based sensor output can be classified into two components: deterministic and random or stochastic. The deterministic component, known as thermal sensitivity, results from the combined effect of thermal expansion strains, resonator loading, Young's sensitivity modulus of silicon (-100 ppm/°C), and changing of the resonator stiffness. The axial load on a resonator element by temperature change is given by [Remtema and Lin, 2000]:

$$P = [\alpha_{si} (T_{si} - T_o) - \alpha_{sub} (T_{sub} - T_o)] \left(\frac{L_s}{L} \right) EA \quad (6.1)$$

where,

P is the axial load caused by temperature change on the resonator,

α_{si} is the temperature coefficient of expansion of silicon,

α_{sub} is the temperature coefficient of expansion of substrate,

T_{si} is the resonator temperature,

T_{sub} is the substrate temperature,

T_o is the initial temperature,

E is Young's modulus of silicon,

A is the resonator beam cross sectional area,

L is the resonator length, and

L_s is the resonator mounting spacing.

However, deformations due to thermal expansion can additionally lead to non-uniform changes in stiffness, resulting in mismatches in the operational mode resonant frequencies.

On the other hand, the stochastic component due to temperature change is known as thermal noise. The thermal noise equivalent acceleration can be expressed as [Fedder, 1994]:

$$\ddot{x} = \sqrt{\frac{4K_b T \omega_o}{m Q}} \quad (6.2)$$

where K_b is Boltzmann's constant, T is the ambient temperature, and ω_o is the resonant frequency of the proof mass m with quality factor Q . Thus thermal noise decreases with increasing mass and quality factor. One further observation of interest can be made by noting that:

$$\frac{\omega_o}{m Q} = \frac{b}{m^2} \quad \text{since} \quad Q = \frac{\omega_o m}{b} \quad (6.3)$$

where the parameter b is the damping coefficient. Substituting in the thermal noise acceleration equation yields:

$$\ddot{x} = \sqrt{\frac{4K_b T b}{m^2}} \quad (6.4)$$

Based on the latter equation, thermal noise increases with an increase in damping [Shkel et al., 1999].

In general, thermal effect is treated as a stochastic error. Nevertheless, it usually contains a deterministic component due to the thermal variation of the sensor's bias and scale factor. In fact, the actual value of the bias and scale factor varies from that obtained through the calibration process, due to the difference between the operational and calibration temperatures. In inertial navigation applications, the thermal variation of both accelerometer and gyroscope biases, if not modeled and compensated, will result in positioning errors that accumulate rapidly with time as shown in the following equations [Shcheglov et al., 2000; El-Sheimy, 2004]:

$$\delta p = \int \delta v dt = \int \int \delta b_f(T) t dt = \frac{1}{2} \delta b_f(T) t^2 \quad (6.5)$$

$$\delta p = \int \delta v dt = \int \int \frac{1}{2} \delta b_w(T) g t^2 dt = \frac{1}{6} \delta b_w(T) g t^3 \quad (6.6)$$

where $\delta b_f(T)$ & $\delta b_w(T)$ are the uncompensated thermal errors of the accelerometer and gyroscope bias respectively, and are functions of the operating temperature (T) in °C, t is the time, g is the local gravity value, δv is the velocity error and δp is the position error.

Practically, several thermally controlled calibration tests are performed utilizing an appropriate climatic chamber to determine the bias and scale factor of the sensor under various temperature points throughout the whole expected operating thermal range [Salychev, 1998]. Results of these tests are then analyzed to develop reliable algorithms describing the thermal variation of calibration parameters of the sensor. Nevertheless, the drift of MEMS-based inertial sensors, due to thermal variation, is always non-linear and

does not even follow a polynomial of second or higher degree order [Helsel et al., 1994]. Therefore a special approach capable of handling such non-linear problems is required. This requirement can be sufficiently fulfilled by the ANFIS, which offers a channel for identification of complicated systems beyond modeling and has a promising capacity in solving non-linear modeling problems. Hence, the ANFIS-based and self-trained prediction networks are adopted herein to model the temperature drift of MEMS inertial sensors.

6.2 Thermal Testing Approaches

The purpose of thermal testing is to establish the variation of the sensor's performance parameters with temperature. There are two main approaches for thermal testing, namely (1) by allowing the temperature of the sensor to stabilize to a certain desired temperature for a period of time during which measurements are recorded, then repeating the same process at different temperature points, and (2) by allowing a controlled linear increase or decrease of temperature over a given period. The first approach is known as soaking or the temperature step stress test while the second is called thermal ramp test. [Titterton and Weston, 1997]. In this thesis, the first approach is implemented at various temperature points to cover the full operational temperature range of the tested sensors. Using this technique, the bias value has been evaluated at each testing temperature point. Consequently, the change of the bias with temperature variation can be observed and either mathematically or graphically expressed then stored in a microprocessor to provide real-time compensation for temperature variations during navigation.

6.3 Results of Lab Thermal Testing of Various MEMS Inertial Sensors

In order to study and verify the effect of temperature variation on the output signal of MEMS inertial sensors, different linear and angular rate MEMS sensors have to be thermally tested. The applied thermal range during the test should include positive and negative temperature ranges that the sensor is anticipated to experience in a real environment. The Attitude and Heading Reference System AHRS400CA from Crossbow (see Figure 6.3) as well as the Systron Donner's MotionPakII (see Figure 6.4) were used for thermal testing as samples of MEMS-based IMUs, utilizing surface micromachining and capacitive technologies. Each IMU integrates six MEMS inertial sensors (three accelerometers and three gyros). Table 6.1 lists some of the important technical specifications for the AHRS400CA IMU while Table 6.2 lists the same for the MotionPakII IMU.



Figure 6.3 Cross-Bow AHRS400CA MEMS-based IMU



Figure 6.4 Systron Donner's MotionPakII MEMS-based IMU

Table 6.1 Technical specifications of the AHRS400CA MEMS-based IMU

	Angular Rate Sensors	Acceleration sensors
Input Range	± 100 ($^{\circ}$ /sec)	± 2.0 (g)
Bias	$< \pm 1.0$ ($^{\circ}$ /sec)	$< \pm 8.5$ (mg)
Scale Factor accuracy	< 1 %	< 1 %
Non-Linearity	< 0.3 (% FS)	< 1 (% FS)
Resolution	< 0.025 ($^{\circ}$ /sec)	< 0.25 (mg rms)
Bandwidth	> 25 Hz	> 10 Hz
Random Walk	< 2.25 ($^{\circ}$ /hr $^{1/2}$)	< 0.1 (m/s/hr $^{1/2}$)
Operating Temperature	-40 to +71 ($^{\circ}$ C)	-40 to +71 ($^{\circ}$ C)
Storage Temperature	-55 to +85 ($^{\circ}$ C)	-55 to +85 ($^{\circ}$ C)

Table 6.2 Technical specifications of the MotionPakII MEMS-based IMU

	Angular Rate Sensors	Acceleration sensors
Input Range	± 75 ($^{\circ}/\text{sec}$)	± 1.5 (g)
Bias	$< \pm 5.0$ ($^{\circ}/\text{sec}$)	$< \pm 125$ (mg)
Scale Factor Error	$< 6\%$	$< 5\%$
Non-Linearity	< 3 (% FS)	< 3 (% FS)
Bandwidth	> 30 Hz	> 20 Hz
Output Noise Analog	6 to 28 mV RMS	5 to 28 mV RMS
Operating Temperature	-40 to +85 ($^{\circ}\text{C}$)	-40 to +85 ($^{\circ}\text{C}$)
Storage Temperature	-40 to +85 ($^{\circ}\text{C}$)	-40 to +85 ($^{\circ}\text{C}$)

6.3.1 Test Description

The two IMU units were mounted inside a climate-controlled chamber SK-2010 (see Figure 6.5) where a soaking temperature test was performed. The temperature step stress began at -30°C and increased in 5°C increments till it reached $+70^{\circ}\text{C}$. As a result, a total of 21 different temperature points were considered during each test. The dwell time at each thermal increment was 20 minutes. The dwell time began after the sensor temperatures, provided by the internal thermal sensor of each IMU unit, stabilized. During the 20 minute dwell time, both IMU output measurements were recorded for 10 minutes with the axis under consideration (e.g. Z-axis) pointing vertically upward and the other 10 minutes with the same axis pointing downward. The same process was repeated for all three axes. Finally, a total of 42 sets of data (21 upward and 21 downward), per

axis, were collected through a single complete thermal test covering the entire operating thermal range of the tested IMU.

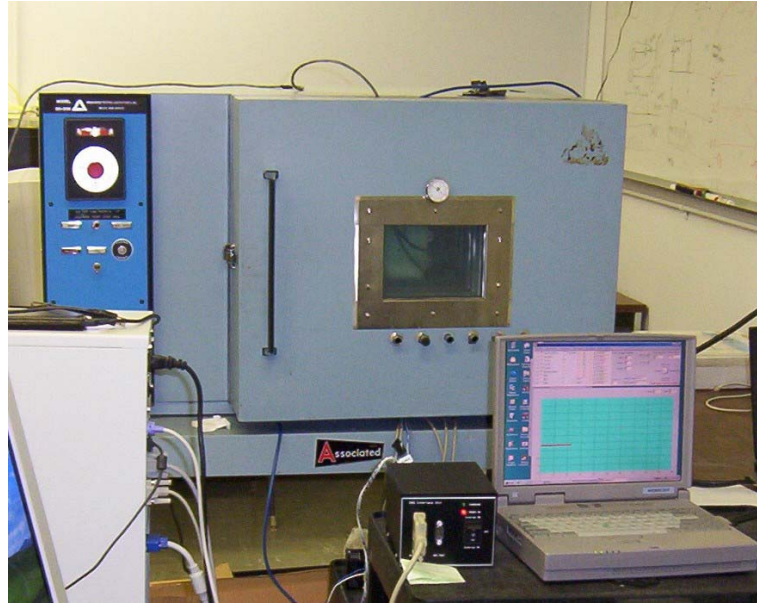


Figure 6.5 SK-2101 Temperature chamber

The recorded data sets were used to estimate the accelerometer bias as well as the gyro drift for each of the three IMU body axes (X, Y and Z) using the following equations [Salychev, 1998]:

$$b = \frac{(f^{up})_m - (f^{down})_m}{2} \quad (6.7)$$

$$dr = \frac{(\omega^{up})_m - (\omega^{down})_m}{2} \quad (6.8)$$

where $(f^{up})_m$ & $(f^{down})_m$ are the mean values of the accelerometer raw output in the upward and downward directions, $(\omega^{up})_m$ & $(\omega^{down})_m$ are the mean values of the gyro raw

output in the upward and downward directions, b is the accelerometer bias, and dr is the gyro drift.

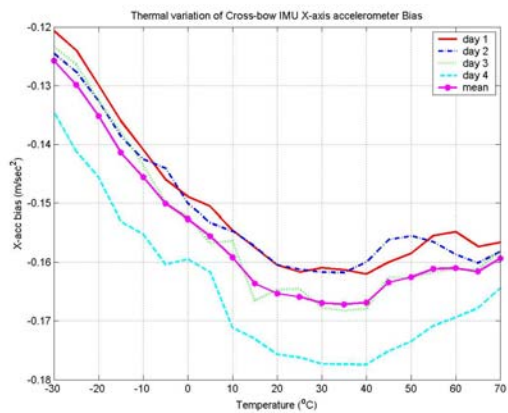
Given the poor run-to-run stability of these MEMS-based sensors, the estimated values of the bias at the various testing temperature points will not be very reliable since it is always contaminated with the so called run-to-run bias. The run-to-run bias, by definition, is a random constant value that changes only from sensor turn on to another [Salychev et al., 2000]. A simple way to overcome this problem is to repeat the above-mentioned thermal test as many times as possible, with the IMU turned off between successive thermal tests. Then, the average bias value of all performed thermal tests, at each desired temperature, can be determined as the best bias estimate after suppressing the effect of the random run-to-run bias instability. Of course the more thermal tests that can be executed, the higher the opportunity to suppress the random effect of run-to-run bias and therefore the more appropriate the bias thermal variation estimate will be achieved. In our case herein, 20 thermal tests were performed for each of the six MEMS inertial sensors included in both IMUs tested.

6.3.2 Thermal Variation of Sensors Calibration Parameters

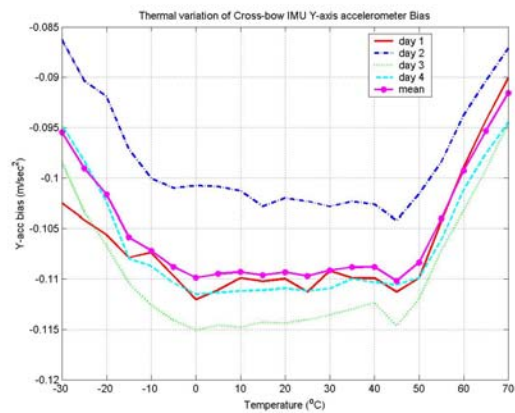
The results of the implemented thermal testing of both IMUs (Crossbow and MPakII) are presented in this section. The main focus will be on the horizontal axes (X & Y) accelerometers and gyros output. It is worthwhile to mention here that the thermal behaviour (trend) of each sensor was almost repeatable throughout all the twenty tests but with slight differences in values, from test to another, within the expected range of run-

to-run random biases of these low cost sensors. Therefore, and for easy plots interpretation, typical results of only four thermal tests are presented in each of the following six figures (Figure 6.6a&b to Figure 6.11a&b). Figure 6.6a&b and Figure 6.7a&b show the variation of bias and scale factor with temperature changes for the Crossbow X and Y-axis accelerometers while Figure 6.8a&b and Figure 6.9a&b demonstrate the same for the MPakII accelerometers. Figure 6.10a&b illustrate the thermal variation of bias (drift) of the Crossbow X and Y-axis gyroscopes while Figure 6.11a&b represent it for the MPakII gyroscopes.

Given that the main intent of this thesis is to enhance the performance of MEMS sensors for its implementation in land vehicle navigations, this work tries to minimize the inertial sensors' biases, particularly gyroscopes', since they are known as the dominant factors that affect the accuracy of the estimated navigation states [El-Sheimy, 2004]. Therefore, the main focus of the following discussion of the results is on the bias thermal variation and its effect on navigation, particularly on the position component.

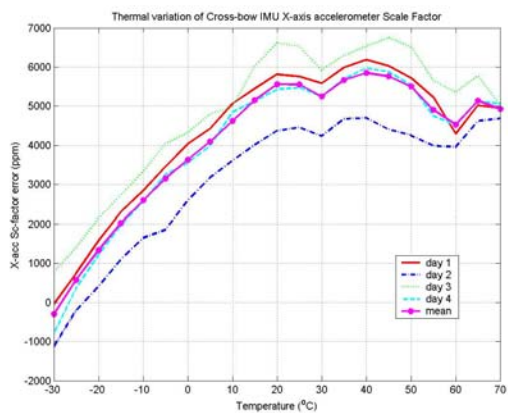


(a)

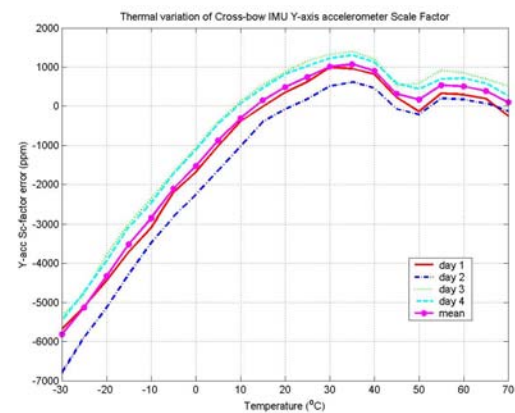


(b)

Figure 6.6 Thermal variation of the Crossbow AHRS400CA X & Y-axis accelerometer biases

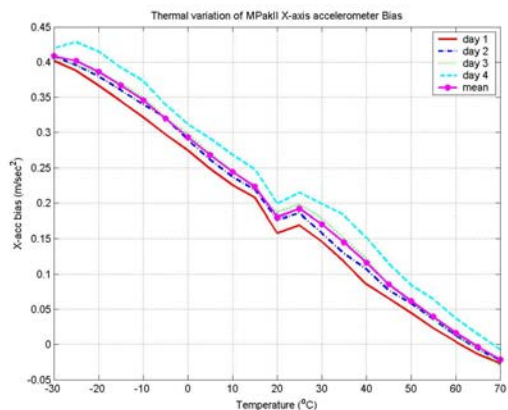


(a)

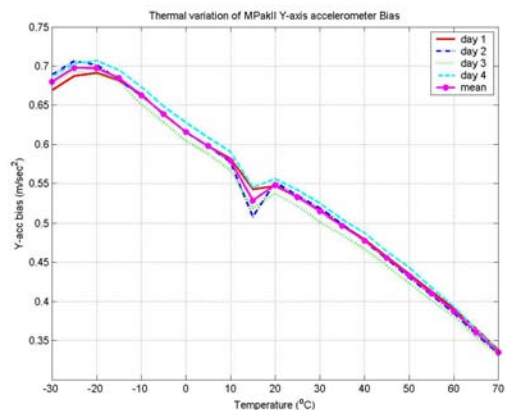


(b)

Figure 6.7 Thermal variation of the Crossbow AHRS400CA X & Y-axis accelerometer scale factors

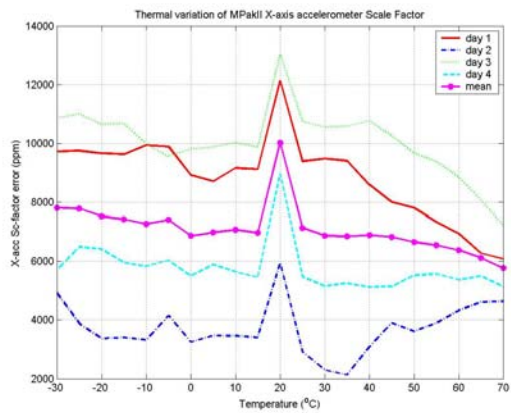


(a)

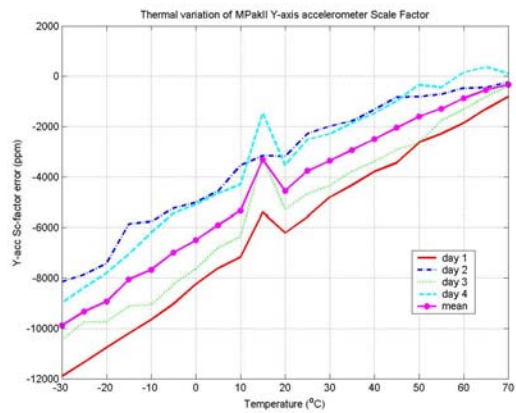


(b)

Figure 6.8 Thermal variation of Systron Donner’s MPakII X & Y-axis accelerometer biases



(a)



(b)

Figure 6.9 Thermal variation of Systron Donner’s MPakII X & Y-axis accelerometer scale factors

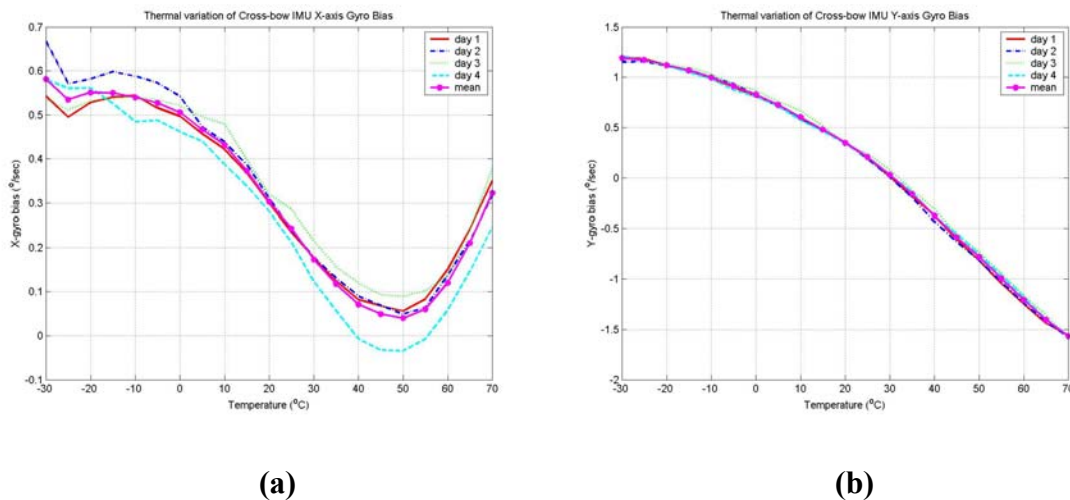


Figure 6.10 Thermal variation of the Crossbow AHRS400CA X & Y-axis gyroscope biases

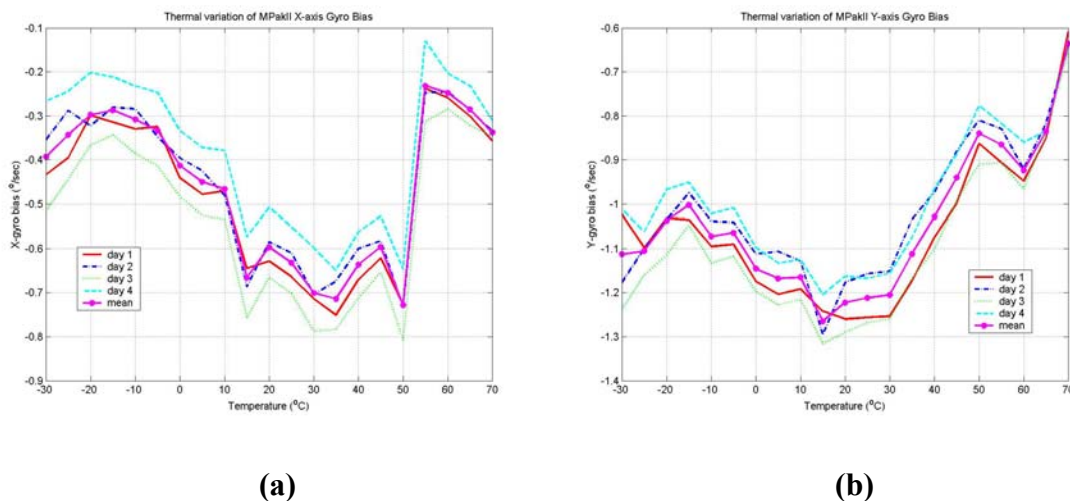


Figure 6.11 Thermal variation of the Systron Donner's MPakII X & Y-axis gyroscope biases

Figure 6.8 shows that the variation of the accelerometer bias over the operating thermal range may reach about 0.3 m/s^2 in case of MPakII accelerometers. On the other hand, the

Crossbow accelerometers are more stable with temperature changes as their thermal bias variation does not exceed 0.05 m/s^2 (see Figure 6.6). Considering the worst case for the MPakII and based on Equation 6.5, if such a thermal error (0.3 m/s^2) if not corrected or compensated for, it can lead to position errors of about 100 m after 30 seconds and about 500 m after one minute of standalone INS inertial navigation. On the other hand, the thermal variation of the gyroscope drift has an average of 0.5 deg/s for both Crossbow and MPakII gyros. Such a gyro error, according to Equation 6.6, may lead to position errors of about 400 m after 30 seconds and almost 3 km after one minute of standalone inertial navigation. These numerical values stress the fact that the gyro drift is the dominant factor when investigating the obtained positioning accuracy during standalone inertial navigation (i.e. without any external update information such as GPS position and velocity updates) [El-Sheimy 2002]. Therefore, in the following sections, the focus will be on the modeling of the thermal variation of gyro drift rather than the accelerometer bias. However, the proposed modeling approach, which is further discussed later on, can be used for the thermal modeling of both sensors (accelerometers and gyroscopes).

The figures clearly show a repeatable non-linear thermal trend for both the accelerometers and the gyroscopes in each of the IMUs. In addition to the non-linearity, the thermal trends are unstable and contain sudden peaks that are difficult to justify from a physical or technology point of view. As a result of this complexity the ANFIS was adopted in this research to aid in solving these non-linear modeling problems which can not be solved using second or higher degree polynomials.

6.4 ANFIS Prediction of Thermal Drift

As discussed earlier in Chapter five, Fuzzy logic has the key advantage of being able to describe the imprecise and non-linear systems' behaviour with simple heuristics or IF-THEN rules. Hence, the Fuzzy networks-based and self-trained ANFIS method (see Section 5.3) is adopted herein to model the unstable and non-linear temperature drift trend of MEMS-based inertial sensors. In the following subsections, the initial FIS structures of ANFIS are presented; the results of its implementation for modeling thermal behaviour of MEMS inertial sensors will then be illustrated and analyzed.

6.4.1 Structure of the Implemented FIS

The general architecture of the ANFIS network that is used for modeling bias thermal variation is illustrated in Figure 6.12. The structure of the FIS that is utilized for ANFIS learning is very simple since it only utilizes a single-input/single-output pattern. Temperature is the only input in this case and the corresponding bias or drift at this temperature is the desired output to predict or estimate. In the applied ANFIS structure, the user only defines the number of inputs and outputs, as well as the number of membership functions (MFs) or clusters for each utilized input. However, the system can help define the number of MFs based on initial analysis of the input/output data sets. During the training process, the system tunes the MFs parameters, training step size, as well as the step decrease and increase rates in order to achieve the optimal modeling algorithm [Jang et al. 1997]. It is pertinent to say that a specific ANFIS network is developed for each sensor included in the IMU, which means a total of six ANFIS

networks are needed to completely describe the thermal trend of each MEMS-based IMU.

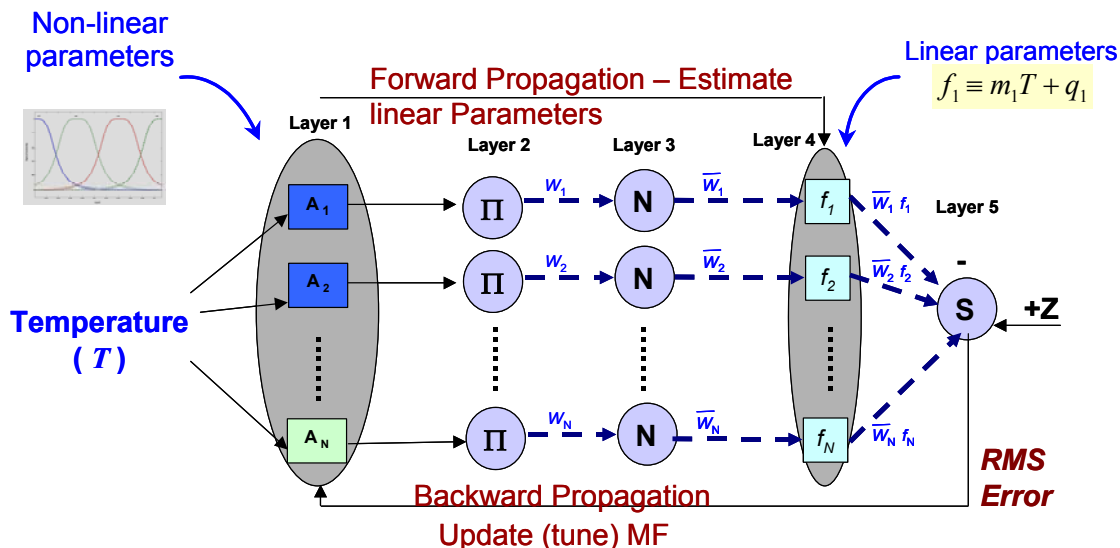


Figure 6.12 Single inputs/single output FIS structure for ANFIS learning

The data sets of the 20 performed tests were divided into two parts. The first part includes two-thirds of the data (13 tests) was used for the actual training. The second part which includes the remaining seven tests data was used for validation and checking. This is a simple and commonly used approach in order to develop the necessary checking data sets, whenever the number of available input/output data points is sufficient [Mathwork's, 2002]. The use of checking sets in addition to the training set is a recommended technique to guarantee ANFIS model generalization and to avoid over fitting of the model to the training data set. Thus, the optimal ANFIS network's parameters obtained at the end of training process are those giving the minimum

estimation error with the checking data. This error is usually called the checking error. Based on the above data splitting strategy, a total of 273 data points (21 temperature points per each of the thirteen tests) were available for training. The optimal number of MFs to be assigned for the single temperature input, in the ANFIS thermal model of each sensor, was iteratively estimated. This optimal number of MFs is different from one sensor thermal model to another. For instance, the optimal number of MFs utilized by the ANFIS-based thermal model of Crossbow X-gyro as well as MPakII X-gyro was estimated to be six. This can be attributed to the high non-linearity of the thermal trend as can be determined through inspecting the previously illustrated gyro drift thermal variations (see Figure 6.10 & Figure 6.11). Using six MFs with a single fuzzy input (temperature) means that six rules are utilized for the appropriate description of thermal trend. This consequently indicates that the attained global ANFIS networks' solution is equivalent to a six polynomial set of equations, which are needed to properly describe the gyro thermal trend in traditional mathematical modeling approaches. It is worthwhile to mention that more than six MFs were iterated (7, 8...etc.) but no significant improvement in the accuracy of the estimated thermal effect was achieved. Conversely, the required time to achieve the ANFIS model convergence was considerably increased. In this regard, the capability of the proposed ANFIS modeling to handle highly non-linear problems with minimal computation load is underlined.

6.4.2 Results of Applying The Proposed Technique

Since the gyro drift is the dominant factor of the standalone inertial navigation accuracy (Equations 6.5 & 6.6), the results of ANFIS modeling of gyro drift is emphasized

hereafter. The results of ANFIS modeling of one typical gyro from each thermally tested IMU (Crossbow and MPakII) is illustrated below followed by a brief analysis and comments. It is worthwhile to mention that similar results and accuracy were obtained for the other two gyros of both MEMS-based IMUs.

Figure 6.13 shows the RMSE of ANFIS prediction for both the training and the checking data as well as the training step size for the training data points. The RMSE represents the root mean square of the difference between the ANFIS predicted outputs and the true outputs calculated during the training stage for all available data points. Figure 6.14 represents the accuracy of the achieved ANFIS algorithm after training. Inputs to the model in this testing process are the same temperature values as used during the training process to cover the anticipated full thermal operating range of the sensor. The output of the ANFIS thermal model at each temperature input is then compared to the mean value of the measured X-gyro drift values, throughout the performed 20 tests, and at the same temperature. The mean value criteria is used herein since it represents the best possible estimate of the gyro drift at the corresponding temperature after removing most of the random run-to-run bias effect. Figure 6.15 shows how the initially selected membership functions of temperature input have been tuned to achieve the optimal modeling of the gyro drift thermal trend. Figure 6.16, Figure 6.17 and Figure 6.18 illustrate the same as the latter three figures but for ANFIS-based thermal modeling of the MPakII X-axis gyro drift.

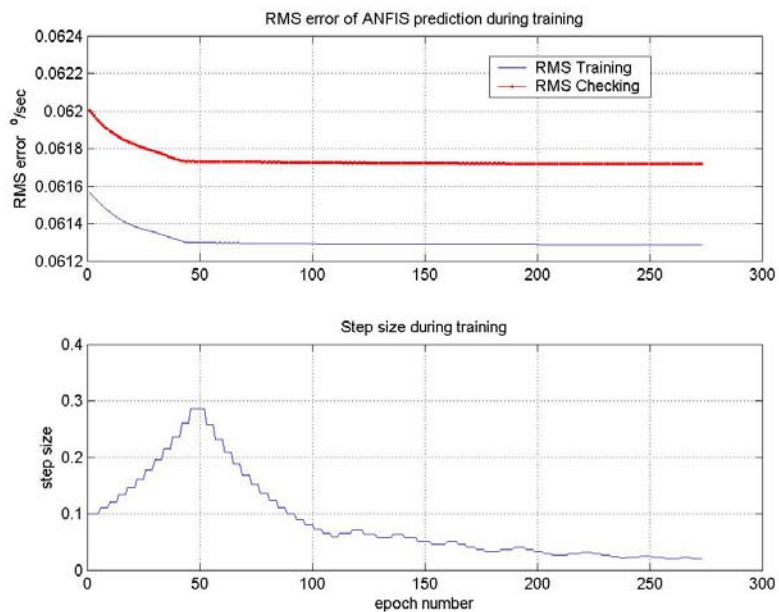


Figure 6.13 RMS error of ANFIS prediction and step size during the training process (Crossbow X-gyro)

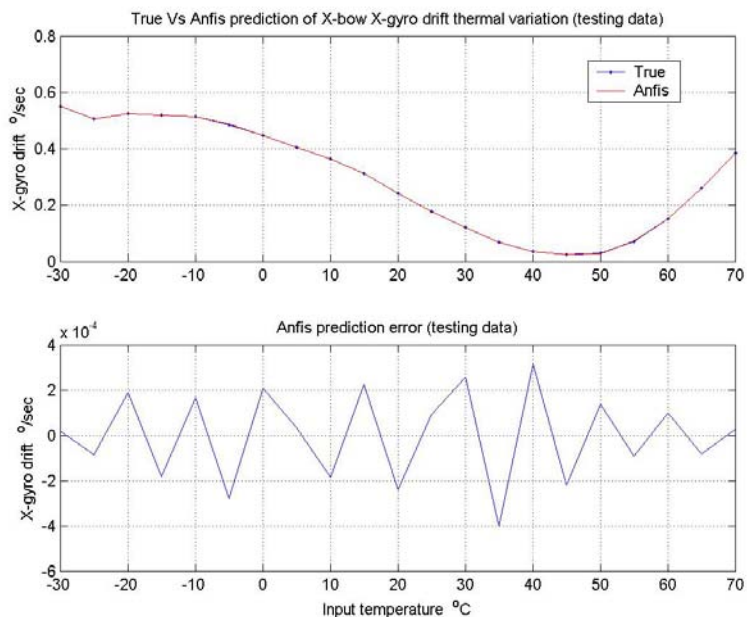


Figure 6.14 Prediction error of developed ANFIS algorithm when applied on test data (Crossbow X-gyro)

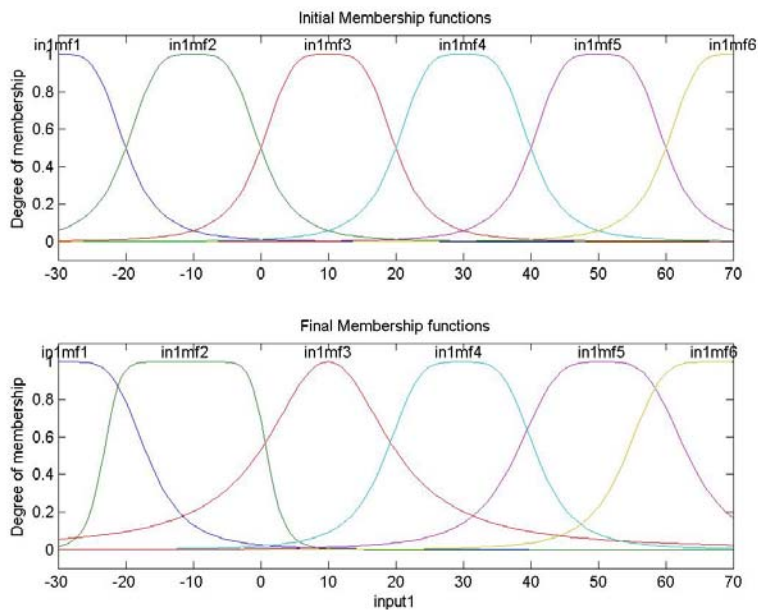


Figure 6.15 Initial and final tuned membership functions of Crossbow X-gyro thermal trend algorithm

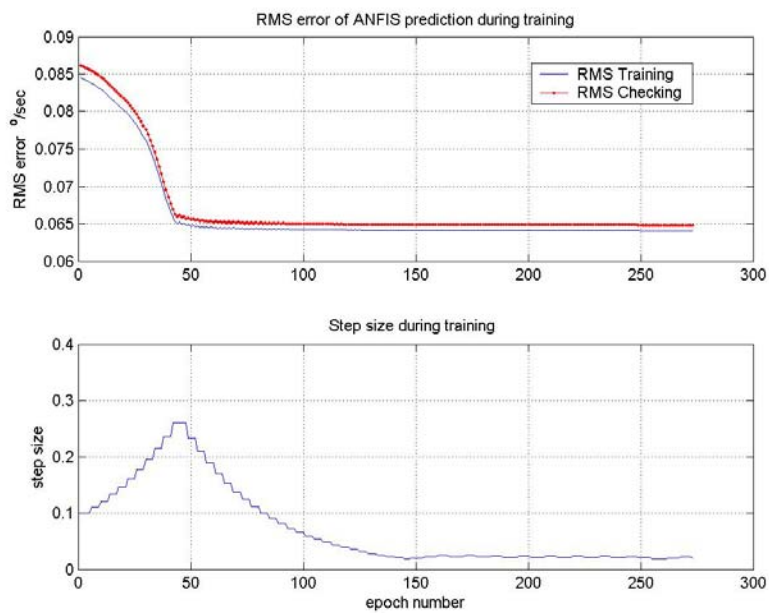


Figure 6.16 RMS error of ANFIS prediction and step size during training process (MPakII X-gyro)

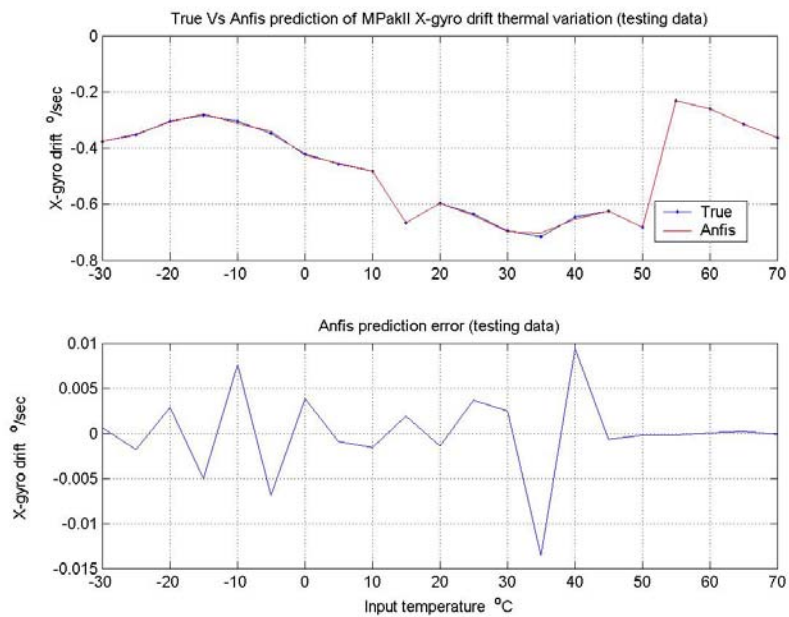


Figure 6.17 Prediction error of developed ANFIS algorithm when applied on test data (MPakII X-gyro)

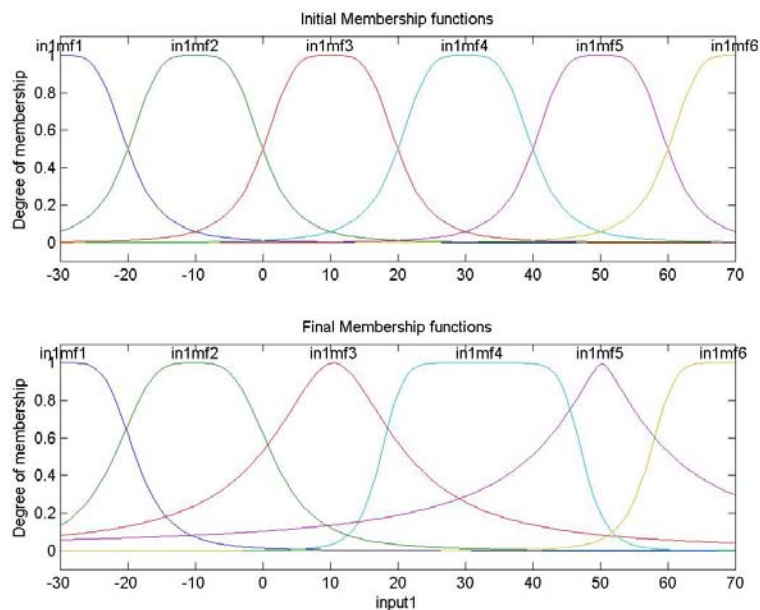


Figure 6.18 Initial and Final tuned membership functions of MPakII X-gyro thermal trend algorithm

6.4.3 Numerical Analysis

Inspection of the training step size and RMS errors of ANFIS prediction (Figure 6.13 & Figure 6.16) reveals how quickly the model converges. As long as the RMS error of the estimated thermal decreases from one epoch to the next, this means that the optimum solution is approached and therefore the training step size is increased. Once the error at a certain epoch is larger than that of the previous epoch, the step size should be decreased with the sign of the non-linear parameters increments reversed. In this way the step size keeps decreasing with the addition of new training epochs till the difference between the errors of two successive epochs is less than a certain pre-specified threshold. At this epoch the optimal solution is achieved, hence the training procedure should be stopped. It can be noticed that the input/output data sets of only two tests (42 epochs) are sufficient to achieve the optimal thermal modeling. The enhancement of the prediction error afterwards, due to considering additional training data sets from other tests, is very slight. This clearly indicates the high repeatability and similarity of the gyro drift thermal trend throughout all the tests. Moreover, the simplicity of the used single-input/single-output FIS helps determine good initial estimates for its linear consequent parameters, even before starting the training process, by using only a couple of samples from each class of the input space. This will definitely result in a small initial RMS error value.

Figure 6.15 & Figure 6.18 show the initial (before training) MFs assigned to the single input (temperature) as well as the resultant tuned MFs (after training). The tuning of the initial MFs parameters is mainly based on a mix of propagation and least squares

algorithms to achieve a minimum RMS error for the predicted gyro drift values with respect to the reference values used from both the training and the checking data sets. In our special case herein where we have only a single input (temperature), the degree of membership of this input directly represents the firing strength of the corresponding rule. Therefore, the resultant tuned MFs indicate the input subspace regions of effectiveness for each of the fuzzy rules implemented in the FIS.

Figure 6.14 and Figure 6.17 clearly indicate how the predicted values of the gyro's drift perfectly match the corresponding reference or true values throughout the entire thermal test range. The maximum error in estimating the gyro drift thermal trend using ANFIS is about 0.005 deg/s. Comparing this value to the expected error ranges discussed in Section 6.3.2, if no thermal corrections were to be considered, it can be deduced that the thermal gyro drift error has been reduced to almost one-hundredth of its original value. This will definitely suppress the expected position error, due to the thermally dependent part of the gyro drift, from hundreds of metres to few metres after 30 seconds and from thousands of metres to tens of metres after 60 seconds of standalone MEMS-based inertial navigation, under such conditions. This clearly indicates the effectiveness of the proposed ANFIS-based technique in predicting the thermal corrections as well as the potential of applying these ANFIS models to the MEMS inertial sensors measurements when being implemented in real time navigation applications.

Based on the previously shown results of ANFIS-based thermal modeling, some important conclusions can be derived. First, the effect of thermal trend of MEMS-based

inertial sensors on the accuracy of standalone inertial navigation parameters, particularly position, has been investigated. Moreover, a new modeling technique, based on ANFIS-networking prediction, has been suggested and assessed to model the high non-linearity of MEMS sensors thermal trends. This is particularly important for the gyro drift since it usually dominates the attainable accuracy of INS standalone navigation. A total of 20 thermal tests were performed on two MEMS-based IMUs (Crossbow AHRS400CA & Systron Donner MPakII) to study the trend and ranges of thermal variations for each of the MEMS linear and angular rate sensors included inside each IMU. The ranges of thermal variations, particularly for gyro drifts were so large (may reach 0.5 deg/s) that it could dramatically degrade the standalone IMU positioning accuracy over time. Given that the thermal error of gyro drift can be reduced to 1/100 of its original value after applying the proposed thermal correction, the anticipated enhancement of standalone MEMS-INS positioning accuracy may reach hundreds of metres (Equation 6.6) over short-term standalone MEMS-based inertial navigations (30 to 60 seconds). The downside is that a considerable effort is needed to calibrate each sensor.

CHAPTER SEVEN

FUZZY PREDICTION OF MEMS-BASED INS/DGPS NAVIGATION ERROR STATES

Performance improvements of integrated IMUs utilizing MEMS sensors, and GPS are presented in this Chapter. The performance improvements are brought about using two new techniques, namely augmentation of the KF and Fuzzy principles. The first proposed technique provides direct corrections to the estimated position by KF during standalone inertial navigation mode (GPS reference solution is not available). The second proposed technique supplies simulated reference positions and velocity error measurements to the KF update engine keeping its functionality during GPS signal outages. Both proposed techniques have been experimentally verified on real MEMS inertial data collected in a land vehicle navigation test. The verification results are illustrated, followed by comprehensive analysis and discussions.

7.1 An Augmented Fuzzy-Kalman For Position Error Prediction

Most of the present INS/GPS integration techniques, particularly in land-vehicular navigation, utilize a conventional KF algorithm. KF is a recursive algorithm designed to compute corrections to a system based on external measurements. In inertial navigation, this can be accomplished by using an external navigation source, such as GPS-derived position and velocity, odometer, Zero Velocity Update (ZUPT), etc. In GPS/INS integration, as long as the GPS position and velocity measurements are available, the KF

solution of INS/GPS integration works efficiently and provides accurate estimates of the navigation states. Nevertheless, during GPS signal outages, standalone INS, specifically MEMS-based, suffers from one serious limitation: errors in the estimated navigation states rapidly accumulate with the passage of time (see Section 3.5). As a result, the overall performance of a MEMS IMU/GPS integrated system is dramatically degraded during GPS signal outages [Abdelazim et al., 2004].

In this section, ANFIS-based networking (see Section 5.3) is proposed as an augmentation to the conventional KF solution in order to overcome the aforementioned problem. The ANFIS model is extracted from a) input/output patterns where the inputs to the fuzzy system are the raw measurements of the MEMS-based IMU, b) the attitude angles determined through mechanization and c) the GPS outage times. The prospective outputs of the fuzzy model are position errors during GPS outages. As long as the GPS signals are available, the difference between the GPS and INS derived positions are used as the desired output for online adaptation of the FIS parameters during the training process. The initial FIS parameters can be extracted first offline using intensive input/output data patterns covering different dynamics and motion scenarios. This proposed augmentation approach has been assessed using data collected from the Mobile Multi-Sensor Research Group MEMS IMU/GPS integrated system in kinematic mode during a van test.

7.1.1 Methodology

The structure of the TS FIS utilized in the ANFIS-based network, which is to augment the KF, is given in the form of its input-output data pattern. In addition, the method of augmenting ANFIS with KF will be discussed in details. The mechanism of the proposed augmented system for MEMS IMU/GPS integration will be analyzed during either the learning or prediction processes. For the sake of simplicity and clarification, some comprehensive schematic diagrams will be also presented.

The compact initial structure and optimal parameters of the core TS-FIS, for the ANFIS-based network, can be extracted through offline training utilizing the Fuzzy-OLS method (see Section 5.4). In order to develop an appropriate initial structure through this offline training, intensive input/output data patterns covering different dynamics and motion scenarios are used. The parameters of this initial ANFIS-based model or network are adapted online by INS sensor measurements as well as part of the mechanization output data, which act as inputs to the ANFIS model. Meanwhile, the difference between KF updated position errors, when GPS is available, and KF predicted ones act as the desired model output during the training process as shown in Figure 7.1.

To achieve the required input/output data entries for ANFIS learning, two KF processing modes should be applied in parallel. In the first mode, the GPS measurements are available at all times and thus the updated error states $X_k^{(+)}$ are always attainable. In the second mode, simulated successive 30-seconds GPS outages are applied and the

predicted error states $X_k'^{(-)}$ (in standalone inertial mode), corresponding to the updated ones $X_k^{(+)}$, are obtained. The difference ($\Delta X_k = X_k'^{(-)} - X_k^{(+)}$) at each epoch is passed to the ANFIS training algorithm as the desired output in addition to the corresponding inputs at the same epochs. In order to perform training in real time navigation, a sliding 30-second length window of input-output data pattern is used for the adaptive optimization of ANFIS antecedent and consequent parameters. Whenever the GPS signals are blocked, for example due to high buildings and overpasses, the ANFIS model switches to prediction mode providing corrections ($-\Delta X_k'$) to the estimated position states calculated by the KF prediction engine, as shown in Figure 7.2. For further clarification, the whole mechanism and functionality of the augmented ANFIS-KF system, in both learning and prediction modes, are illustrated in Figure 7.3 and Figure 7.4 respectively.

The proposed fuzzy model uses 10 states as input. These inputs are the following:

- The time elapsed since losing the GPS signal.
- The three linear accelerations (F_x, F_y, F_z) measured directly by the three MEMS accelerometers of the low-cost INS system.
- The three angular velocities ($\omega_x, \omega_y, \omega_z$) measured directly by the three MEMS gyroscopes of the low-cost INS system.
- The three attitude angles determined in the local level frame as direct output of the mechanization process.

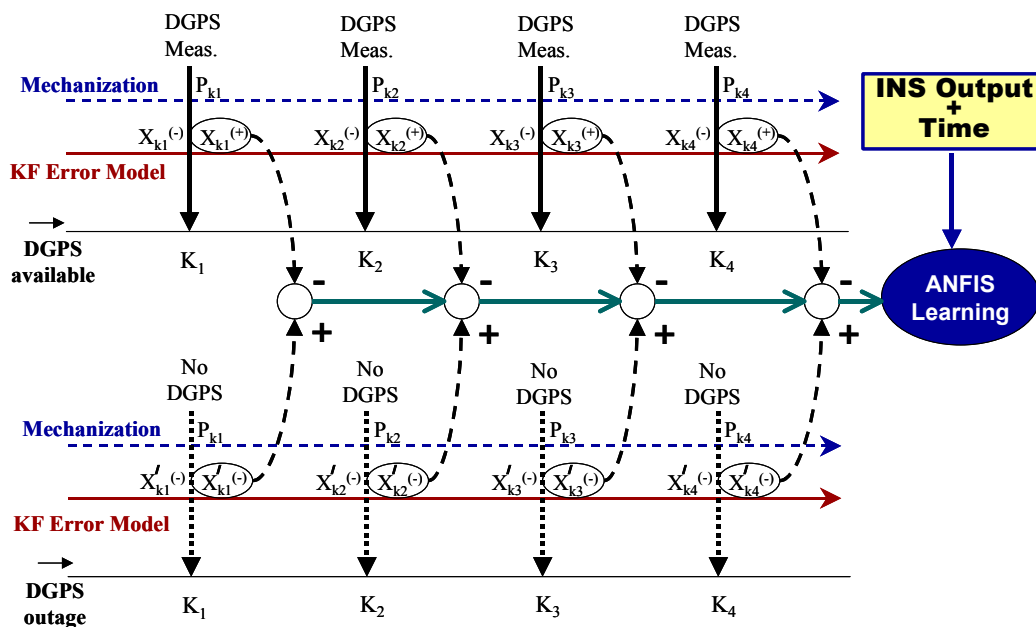


Figure 7.1 Extracting the input/output data entries for ANFIS learning

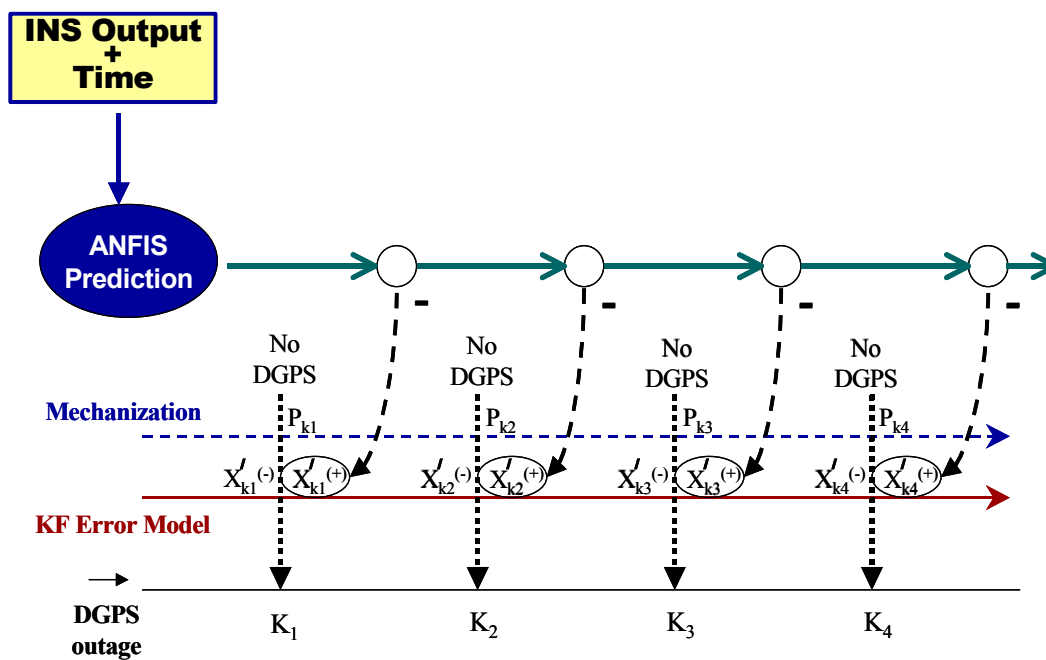


Figure 7.2 Corrections provided by ANFIS to the KF predicted position errors

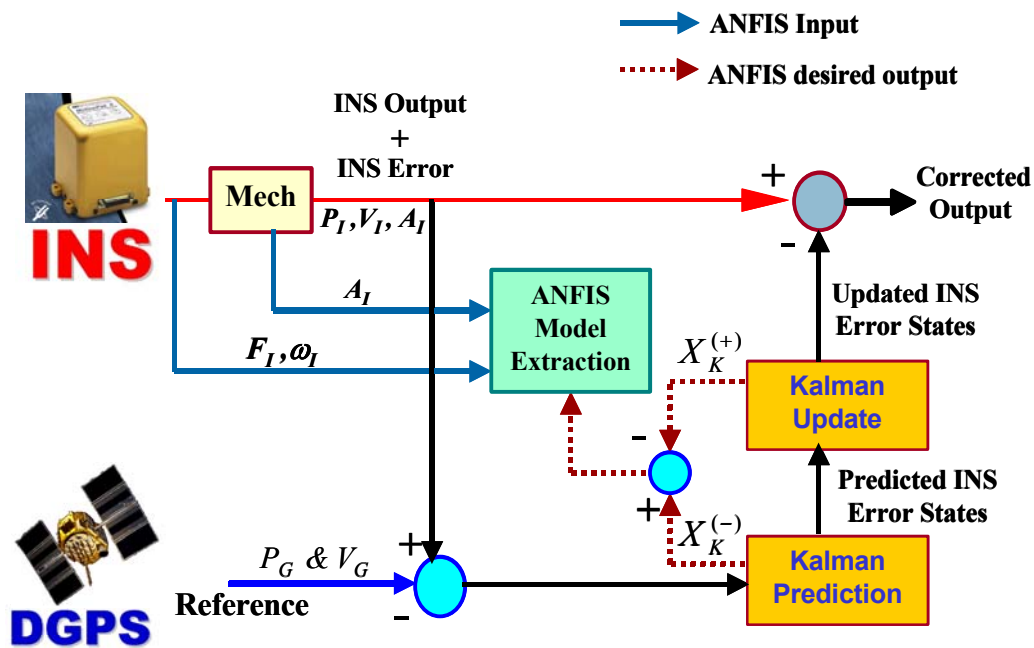


Figure 7.3 Working mechanism of the augmented ANFI-KF system for providing direct position corrections (learning mode)

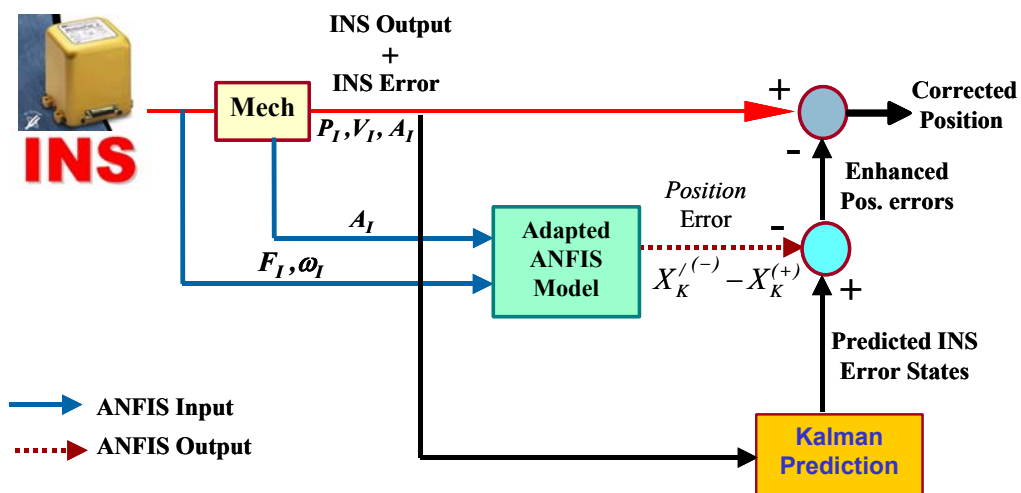


Figure 7.4 Working mechanism of the augmented ANFI-KF system for providing direct position corrections (prediction mode)

The selection of these parameters as inputs to the ANFIS is convenient (Equation 3.7) since they are the main factors affecting the outputs of the Fuzzy model (position error components). Recalling Equation 3.7:

$$\begin{aligned}\delta \dot{r}^{\ell} &= D^{-1} \delta v^{\ell} \\ \delta \dot{v}^{\ell} &= -(2\Omega_{ie}^{\ell} + \Omega_{el}^{\ell}) \times \delta v^{\ell} - \delta R_b^{\ell} f^b + R_b^{\ell} \delta f^b + \delta g^{\ell} \\ \delta \dot{A}^{\ell} &= E \delta v^{\ell} + R_b^{\ell} \delta \omega^b\end{aligned}\quad (7.1)$$

It can be seen that the main sources of position errors are the sensor's (accelerometer and gyros) measurements errors δf^b and $\delta \omega^b$, the attitude and heading angles included in R_b^{ℓ} , as well as the time since losing the GPS signal. Each accelerometer measurement error can be classified as either a bias or a scale-factor error. The first may be considered as a constant and its effect is proportional to (t^2) while the latter is a portion of the acceleration measurement itself. The same can be mentioned about the angular rate measurement error except that the gyro bias effect is proportional to (t^3) [El-Sheimy, 2002]. The sine of the tilt angles (θ_N, θ_E) , which are always small in land-vehicular applications, can be expressed by the values of the angles in radians, while their cosine is very close to zero. On the other hand, and unlike the tilt angles, the heading angle can have a wide range of values $(0$ to $360^\circ)$ and therefore it is preferably expressed by its sine and cosine components.

Fortunately, the adopted inputs are all easy to obtain since parts of them are direct measurements of the inertial sensors while the rest are simply determined through the

INS mechanization process (Equation 3.6). The desired outputs of the ANFIS model are the three position component errors during the absence of the update (reference) position measurements provided by the DGPS solution.

7.1.2 Experimental Validation

In order to examine the effectiveness of the proposed ANFIS model and its impact on the accuracy of its predicted navigation parameters (particularly position), field test data incorporating MEMS-based IMU and DGPS was used. The field test data (provided by Applanix Corporation, Canada) was recorded during a van test involving MEMS-based ISI IMU containing Analog Devices MEMS gyros and Colibrys MEMS accelerometers (Table 7.1 lists the specifications of this IMU). In addition, an Applanix POS-LV-320 INS that uses a tactical grade LTN200 IMU and a NovAtel Millennium GPS receiver was also involved. A double differenced DGPS solution was performed using another NovAtel Millennium receiver at the master station. The average standard deviations (1σ) of DGPS measurements, in kinematic mode, were 1-2 cm for horizontal positions, (3-4 cm) for vertical positions, (0.2-0.25 m/s) for horizontal velocities, and 0.3-0.4 m/s for vertical velocities. Both DGPS and LN200 trajectories were available throughout the entire test and could be used as an accurate reference for the low cost MEMS IMU.

Inertial measurements of MEMS sensors were integrated with the DGPS output and processed through the INS Tool Box [Shin and El-Sheimy, 2002] KF software. Several simulated 30-s GPS outages were introduced at different locations along the trajectory, as shown in Figure 7.5. The test was performed in an open sky where a sufficient number of

satellites were available throughout the whole time of the test which afforded a good satellite geometry and consequently a highly accurate DGPS solution. It is worthwhile to mention that the applied scenario is idealistic as the GPS accuracy is very high to the very epoch where a total outage is simulated. In reality, a gradual degradation of satellite geometry and measurement accuracy due to signal multipath would very likely dramatically affect the results. However, a suggested solution to this problem is to assign an explicit DGPS accuracy below which the online ANFIS parameters update should be stopped. Different dynamics and motion types (straight and ninety degree left-turns) were considered when choosing the time of these simulated GPS outages. The position errors before and after applying the proposed ANFIS model were studied.

Table 7.1 Specifications of the ISI IMU



Dimension

- Size (2.8" x 3.0 " x 2.3"),
- Weight (nominal weight is less than 0.8 lb.)

Technical Specification

Parameter	Accelerometer	Gyro
Range	+/- 20 (g)	+/- 90(deg/s)
Short-term bias stability	+/- 2 (mg)	<0.01 (deg/s)
Scale factor accuracy	0.2%	0.2%
Long-Term bias stability	NIL	< 1 (deg/s)
Non-linearity	0.5% full scale	0.2% full scale
Random Walk	<0.1 ()	<0.5 ()
Bandwidth	>= 30 Hz	>=30 Hz
Data resolution	1 (mg)	2(arcseconds/sec)

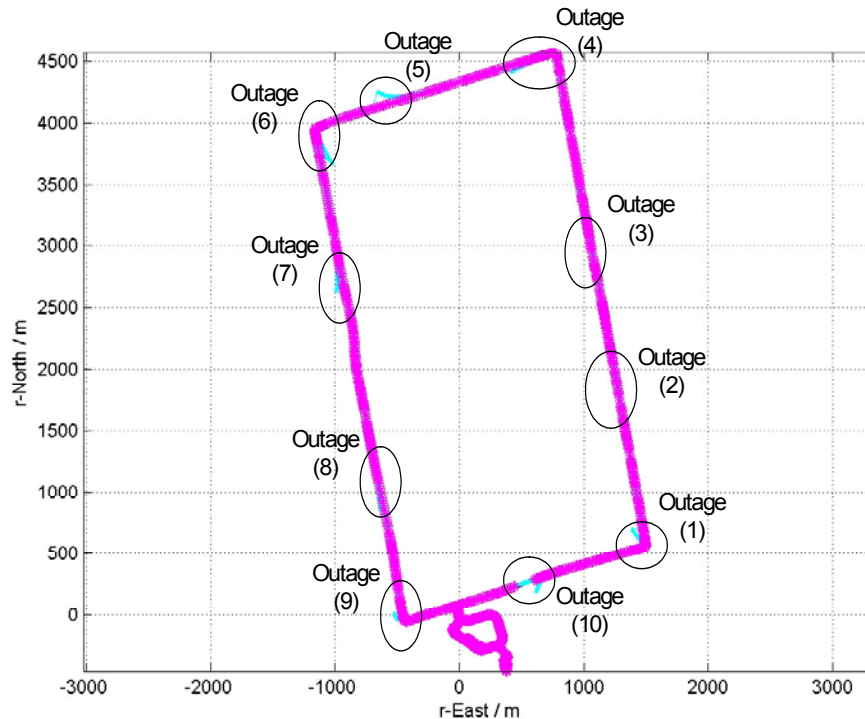


Figure 7.5 Test field trajectory with simulated GPS outages indicated

7.1.3 Results and Analysis

Figure 7.6, Figure 7.7 and Figure 7.8 show the position errors ($\delta\Phi, \delta\lambda, \delta h$), during GPS outages 2, 4 and 7, respectively, as provided by the KF (in prediction mode) against those after applying predicted corrections by the proposed ANFIS algorithm. Vehicle motion was almost straight during outages 2 and 7 while there was a 90° left-turn included during outage 4. Position error drifts are illustrated graphically during these three outages, as three typical examples of vehicle straight motion and manoeuvres, while the results of the remaining outages are only presented numerically (Table 7.3).

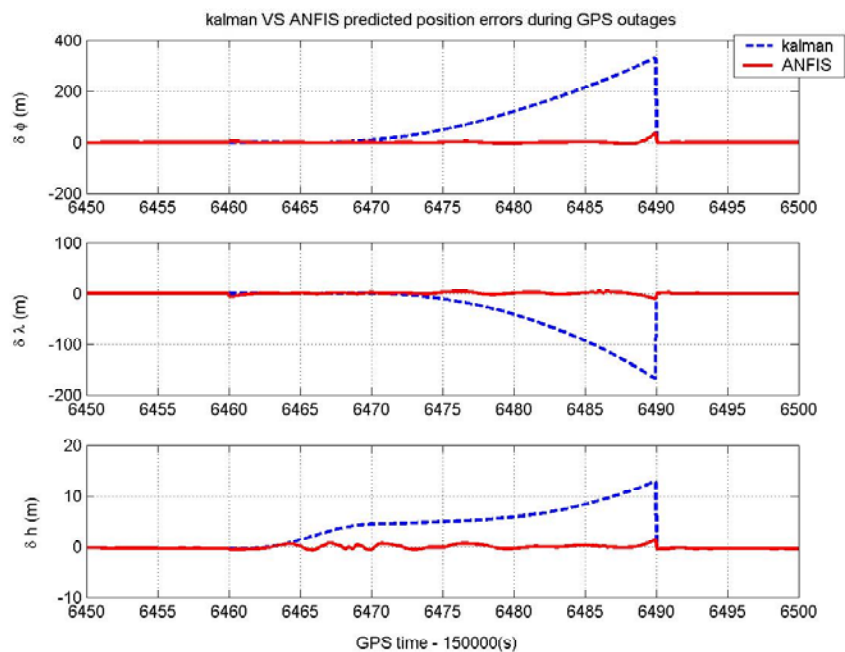


Figure 7.6 Position accuracy during GPS outage (2) before and after ANFIS position corrections

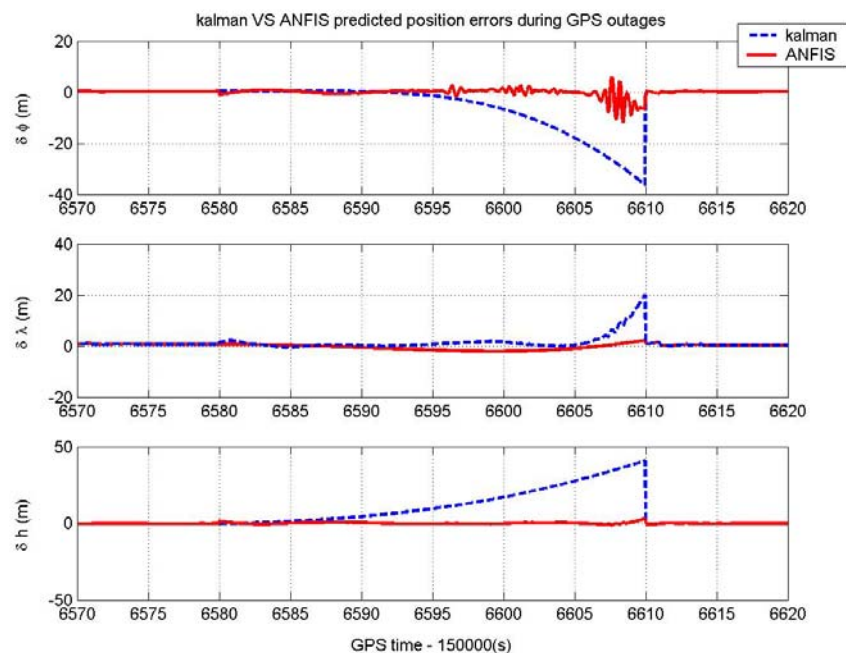


Figure 7.7 Position accuracy during GPS outage (4) before and after ANFIS position corrections

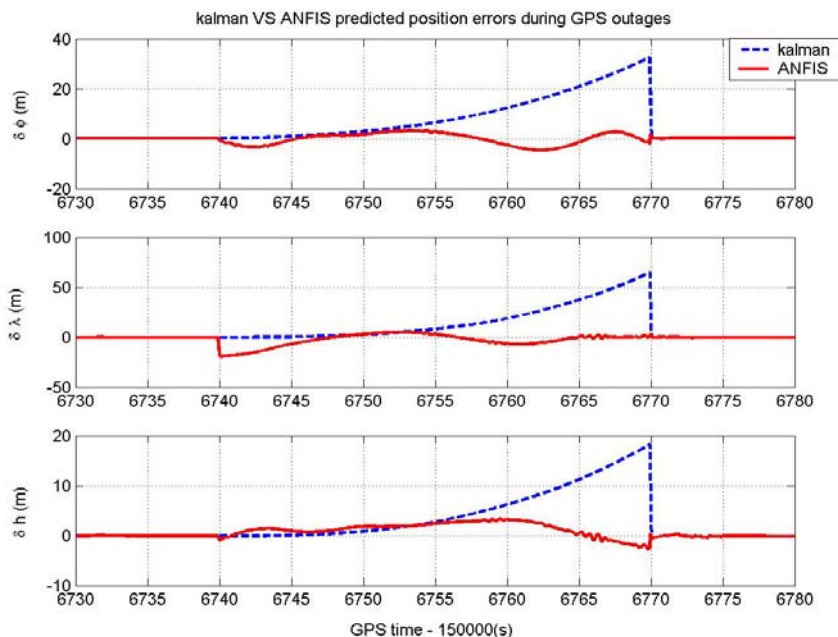


Figure 7.8 Position accuracy during GPS outage (7) before and after ANFIS position corrections

The figures clearly show that significant improvement in the positioning accuracy can be achieved after applying the proposed ANFIS algorithm corrections during these ideal GPS signal outages. Table 7.2 lists the statistics of the positioning accuracy enhancement during the GPS signal outages. The maximum error of each position component, typically achieved at the end of the outage period, was used as a measure of the KF performance before and after applying the ANFIS predicted corrections. Consequently, a rough estimate of positioning accuracy enhancement, as a percentage of the original error, can be obtained during these GPS signal outages. It is worthwhile to mention here that the accuracy of the conventional KF estimated position is always superior over that of the ANFIS estimates for the first few 5 to 10 seconds. This is because the effect of error accumulation with time, dominating the performance of the KF solution, is still minor

during those initial few seconds right after losing the GPS updating measurements. On the other hand, the proposed ANFIS does not experience the same behaviour. and hence the maximum error of the ANFIS estimated position is not accumulated at the end of the GPS signal outage period. Moreover, the maximum error of the ANFIS-based estimated position is likely to happen at the beginning of the GPS outage period, therefore it is recommended to use the KF solution for the first few seconds (less than 10 seconds) of GPS signal outages. For the sake of completeness, the numerical representation of positioning accuracy enhancement, during all of the simulated GPS signal outages, is given in Table 7.3. In this table, the positioning RMSE during each of the 30-second GPS signal outages, after performing the proposed augmented solution was compared to the results obtained from the conventional KF solution.

Table 7.2 Positioning error reduction after applying the augmented solution

Outage Number	GPS Outage (2)	GPS Outage (4)	GPS Outage (7)
Φ error after KF only (m)	330.10	-36.30	32.80
Φ error after ANFIS (m)	36.80	-6.80	2.80
Enhancement (%)	89 %	81 %	91 %
λ error after KF only (m)	-167.20	20.05	65.10
λ error after ANFIS (m)	-11.30	2.33	3.80
Enhancement (%)	93 %	88 %	94 %

Ht error after KF only (m)	12.90	40.75	18.30
Ht error after ANFIS (m)	1.50	3.05	3.20
Accuracy enhancement (%)	88 %	92 %	83 %

Table 7.3 Positioning accuracy enhancement after applying the augmented solution

Position error	Final Error value									
	Outage (1)	Outage (2)	Outage (3)	Outage (4)	Outage (5)	Outage (6)	Outage (7)	Outage (8)	Outage (9)	Outage (10)
KF RMSE Φ (m)	42.0	9.5	4.9	104.8	11.6	12.2	10.5	84.7	39.2	14.1
ANFIS RMSE Φ (m)	6.7	1.6	0.6	3.2	0.5	3.9	1.9	6.9	8.1	2.1
KF RMSE λ (m)	11.3	2.9	45.0	47.0	6.6	27.9	18.8	31.7	11.2	22.7
ANFIS RMSE λ (m)	3.90	1.1	1.5	1.8	0.9	1.5	5.4	8.7	2.6	4.7
KF RMSE h (m)	14.7	13.7	5.1	4.7	5.9	9.2	5.6	9.1	9.2	10.7
ANFIS RMSE h (m)	6.1	0.4	0.4	0.4	0.5	2.9	1.5	1.3	1.3	1.8

Through a careful inspection of the results presented in this section, some important remarks and conclusions can be extracted. The proposed augmented ANFIS-KF model provides reliable position error estimates that can be used to correct the inaccurate results predicted by KF during the absence of the reference DGPS position measurements. In other words, the experimental results demonstrated the ability of the proposed techniques

to reimburse the absence of the update DGPS measurement, under the ideal scenarios used, and keep consistent positioning accuracy. Without the use of the proposed model, the solution of the conventional KF is very large for many land vehicle navigation applications and the system has a degraded performance during GPS outages. Based on the results of the simulated case study and data set used in this thesis, a significant improvement of at least 75-80% in the overall positioning accuracy can be obtained after applying the proposed augmented model. Although these initial results cannot be generalized, they may be promising enough to give the green light for further future research considering real GPS outages and different motion scenarios for the sake of generalization.

7.2 An Augmented Fuzzy-Kalman to Preserve Update Engine Functionality

In this section, an augmented ANFIS-Kalman solution, similar to that presented in the previous section, is adopted. The main objective this time, however, is to preserve the functionality of the KF prediction engine by providing compensatory simulated reference position and velocity error measurements during the absence of a DGPS solution. Although the structure of the ANFIS-based network used in this section is the same, the training and prediction strategies are different than the ANFIS-based network used in the previous section. Again, the initial FIS parameters and structure can be primarily developed through offline training using intensive input/output data patterns covering different dynamics and motion scenarios. The method of extracting the input/output data entries as well as the mechanization of the augmented system, in both learning and prediction modes, will be analyzed in the following subsection.

7.2.1 Methodology

Similar to the proposed ANFIS network presented in Section 7.1, the compact initial structure with its antecedent and precedent parameters can be developed through offline training utilizing the Fuzzy-OLS method. The inputs to the TS-FIS also include the INS sensors measurements as well as a subset of the mechanization results. The difference between the KF updated position and velocity errors (after acquiring each GPS measurement) and the KF predicted results (right before DGPS measurement) provides the desired model output during the training process. In this case, there is no need to perform two parallel KF processing modes, like the previously proposed technique in Section 7.1. This is because the values of the Kalman gain (recall Equation 3.14) during GPS availability will directly constitute the vector of desired outputs needed for training. This will definitely results in less computation load and processing requirements.

The difference between predicted and updated position, the velocity error $(X_k^{(+)} - X_k^{(-)})$ at each DGPS measurement epoch, is passed to the ANFIS training algorithm as the desired output in addition to the INS mechanization results at the same epoch (Figure 7.9). In this case, even though the time is still provided as a TS-FIS input, its effect is insignificant similar to the case described in the previous section. This is because the ANFIS training is performed only over the short time between two successive GPS measurements (normally one second), which will guarantee insignificant time-drift position or velocity error. Whenever the GPS signal is blocked, the ANFIS model

switches to prediction mode to provide simulated reference measurements ($Z_k = X_k^{(+)}$) through adding the correction ($X_k^{(+)} - X_k^{(-)}$) to the KF predicted error states $X_k^{(-)}$ (Figure 7.10). In this way, the KF will keep working in update mode even during GPS outages.

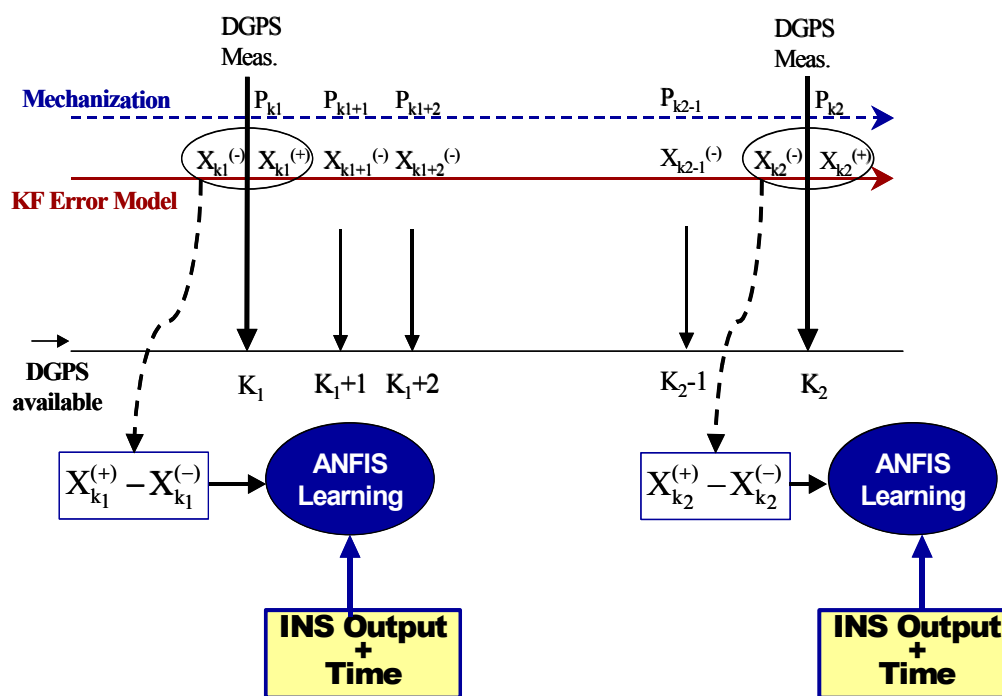


Figure 7.9 Input/output data extraction during ANFIS training

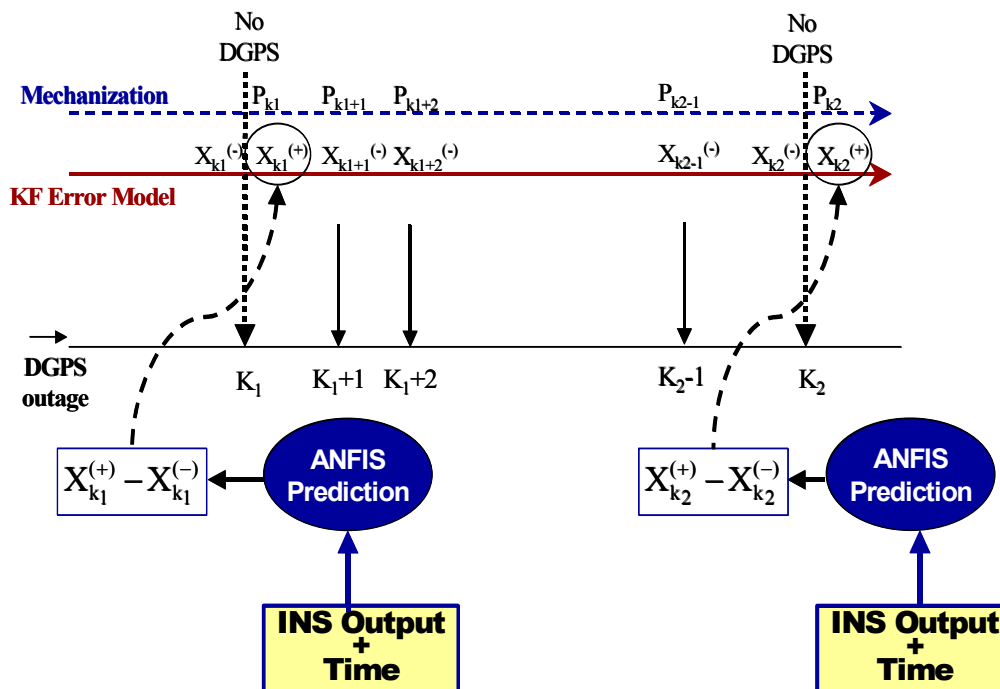


Figure 7.10 Providing simulated reference error measurements by ANFIS

The problem now is how to assign values to the reference measurements covariance matrix in order to determine the Kalman gain. Note that the extracted desired output during training is $(X_k^{(+)} - X_k^{(-)})$, which comprises the effect of the Kalman gain (K_k) and the covariance of DGPS measurements R_k (recall Equations 3.14 and 3.13). This means that the parameters of the TS-FIS are always tuned to predict the actual IMU-derived position or velocity errors after applying the mechanization equations and the KF dynamic error model. Therefore, the simulated error measurements, provided by ANFIS model during GPS outages, are definitely equivalent to the updated error states which necessitates either taking K_k equal to one or assigning R_k zero covariance elements. The mechanism and functionality of the augmented ANFIS-KF system in learning mode is

similar to that given in Figure 7.3, while that in prediction modes is illustrated in Figure 7.11.

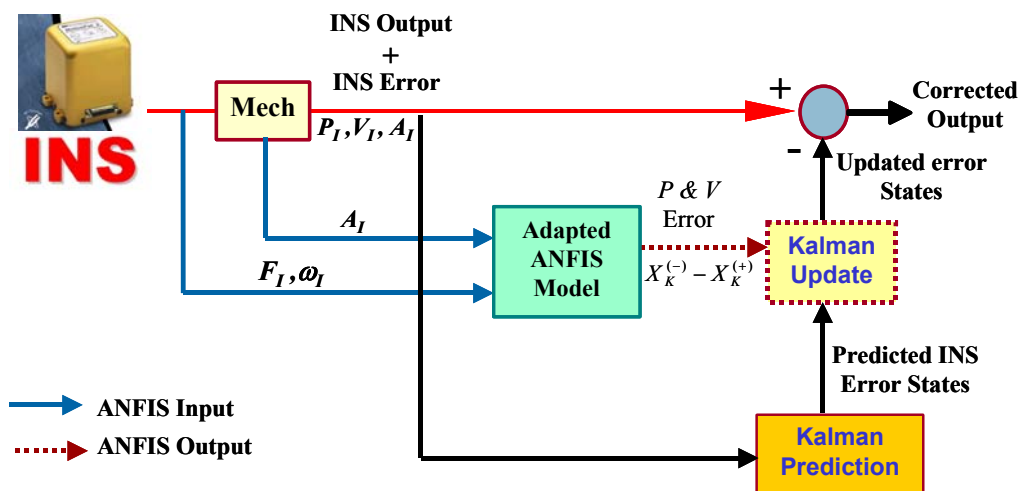


Figure 7.11 Working mechanism of the augmented ANFI-KF system for preserving KF update engine functionality (prediction mode)

7.2.2 Experimental Validation

To test the effectiveness of the proposed augmented ANFIS-KF system, and its impact on the accuracy of the computed navigation parameters (position and velocity), the same field test data (subsection 7.1.2) and its trajectory (Figure 7.5) were used. Also the same INS Tool Box [Shin and El-Sheimy, 2002] was implemented to process the inertial measurements of MEMS sensors with DGPS output. Several simulated thirty-second (30-s) GPS outages were introduced at different locations along the test trajectory (see Figure 7.5). Different dynamics and motion types (straight and 90° left-turns) were considered when choosing the time of these simulated GPS outages. The position and velocity errors

were determined before and after applying the proposed ANFIS-KF technique for comparison and verification.

7.2.3 Results and Analysis

Three simulated GPS outage results are presented. These outage periods are denoted as outage 3 (Time: 5630-5660), outage 5 (Time: 5930-5960) and outage 6 (Time: 5960-5990). During these simulated outages, the KF works in prediction mode and therefore the position and velocity errors are expected to increase with time. On the other hand, the proposed method will work in update mode where the update measurements are obtained from the ANFIS network. Figure 7.12, Figure 7.13 and Figure 7.14 show the velocity drift during the three simulated GPS outages 3, 5, and 6 respectively, before and after applying the proposed ANFIS-KF technique. While, Figure 7.15, Figure 7.16 and Figure 7.17 show the position drift during the same outages. The truth model for these figures was obtained using the trajectory computed from the Applanix POS-LV-320 system.

It can be seen from the above figures that the maximum errors in the position ($\delta\phi$, $\delta\lambda$, δh) as well as velocity errors (δv_E , δv_N , δv_U), which typically happens at the end of the simulated outage, have been successfully reduced. It is worthwhile to mention that similar accuracy enhancements were obtained during the other simulated GPS outages. Table 7.4, Table 7.5 and Table 7.6 summarize the RMSE and the final error value for both position and velocity components for both the KF only and the proposed augmented model at the end of each of the three simulated GPS outages. The results clearly indicate

that a sub-metre positioning accuracy level can be achieved through the application of the integrated ANFIS-KF model under the scenarios tested. These attainable sub-metre positioning accuracy levels might be a little bit worse in other simulated GPS blockage periods particularly along straight lines (i.e. no manoeuvres). In this case, the positioning accuracy might be a bit degraded to reach a few meters. This still can be seen as significant accuracy improvement when compared to the results obtained from KF only which may reaches 120 m in position as in outage 3.

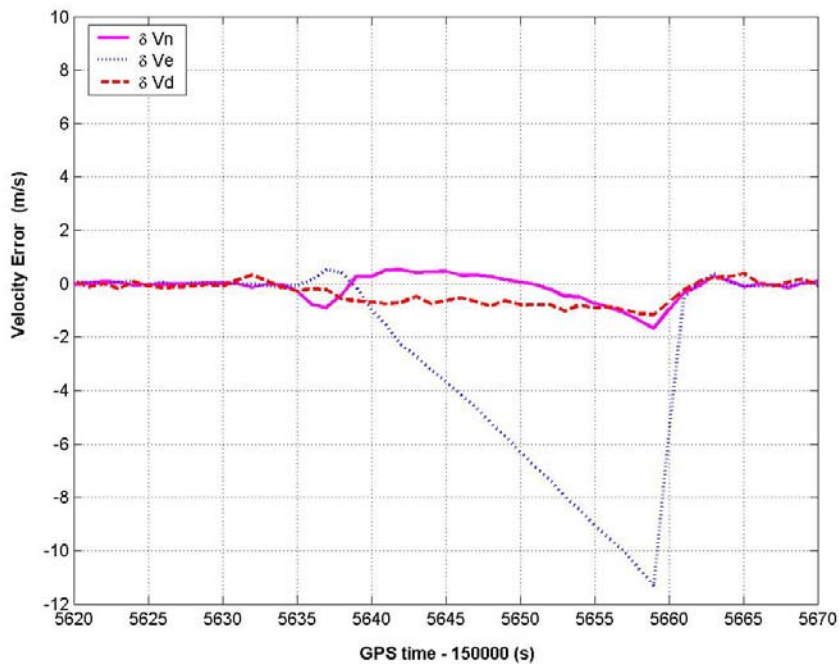


Figure 7.12a Kalman estimated velocity drift during outage (3)

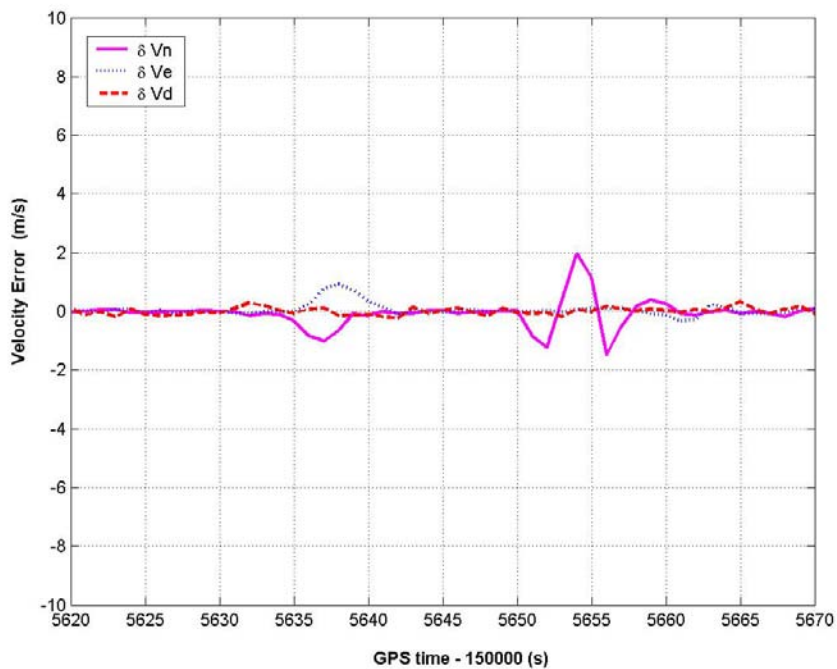


Figure 7.12b ANFIS-KF estimated velocity drift during outage (3)

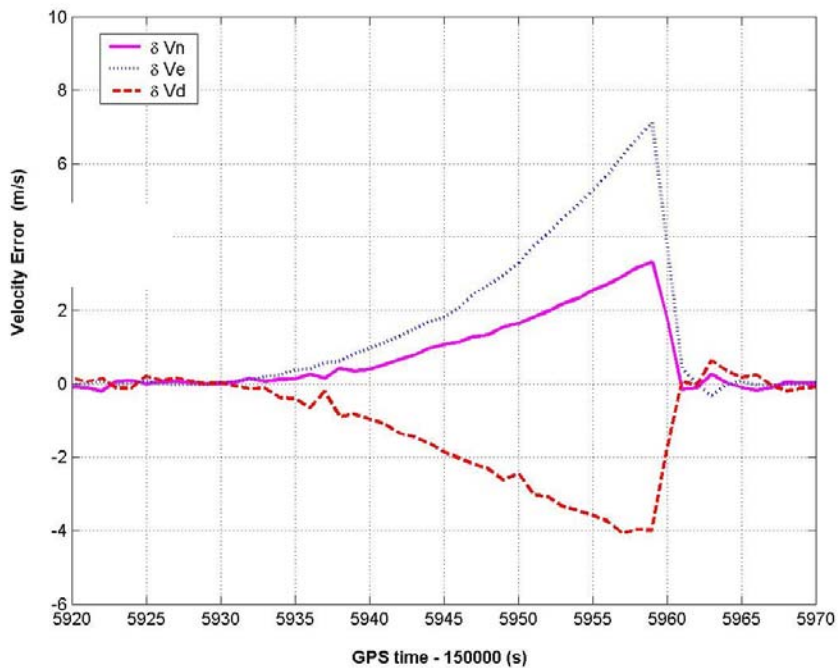


Figure 7.13a Kalman estimated velocity drift during outage (5)

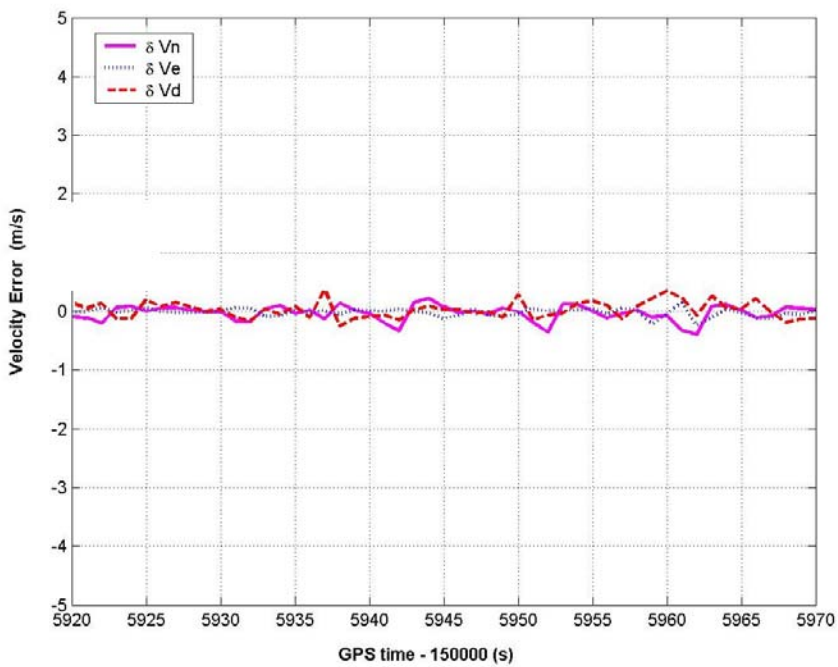


Figure 7.13b ANFIS-KF estimated velocity drift during outage (5)

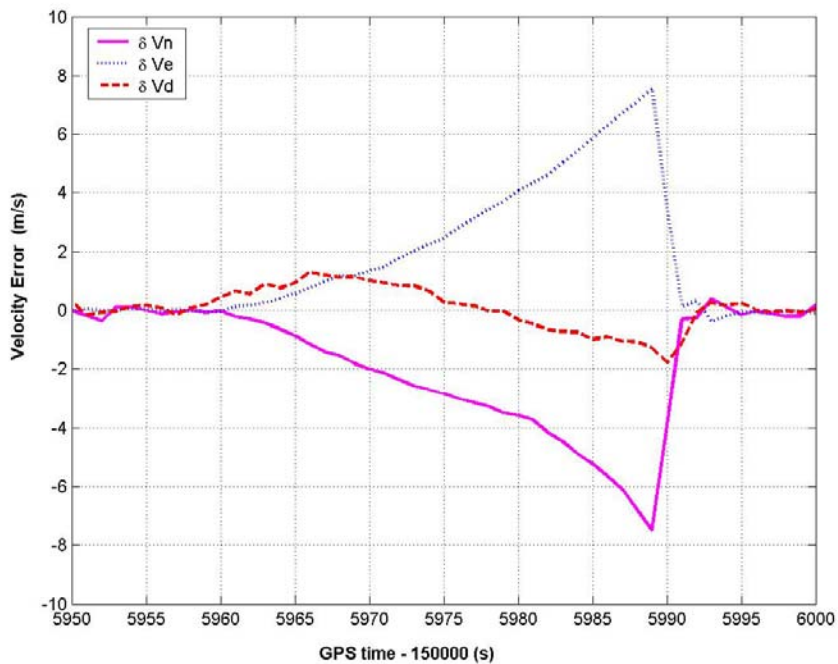


Figure 7.14a Kalman estimated velocity drift during outage (6)

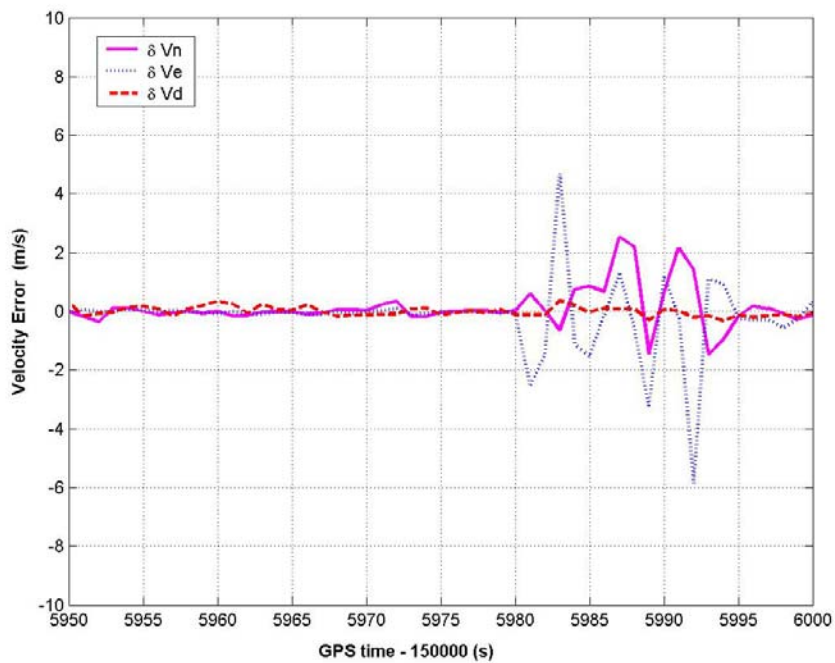


Figure 7.14b ANFIS-KF estimated velocity drift during outage (6)

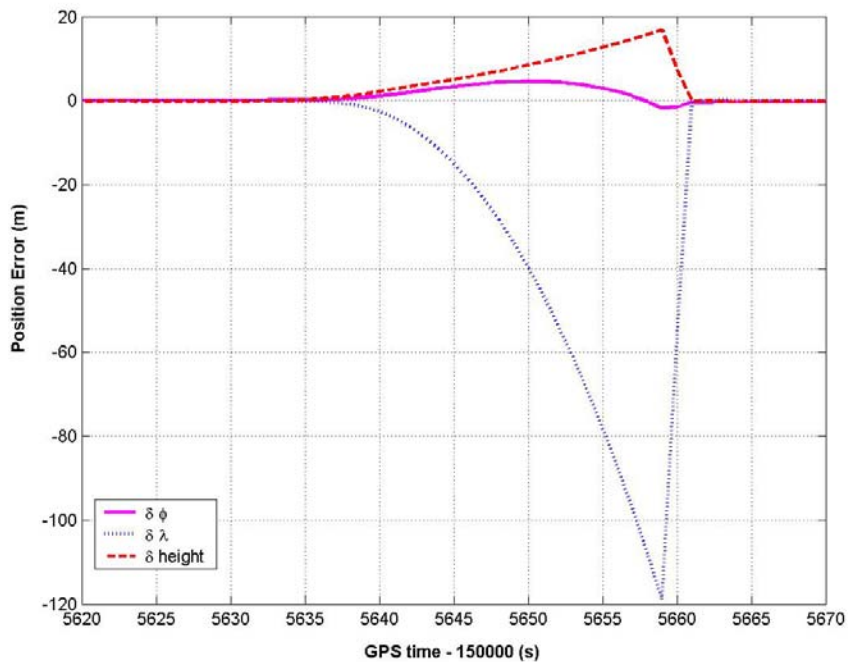


Figure 7.15a Kalman estimated position drift during outage (3)

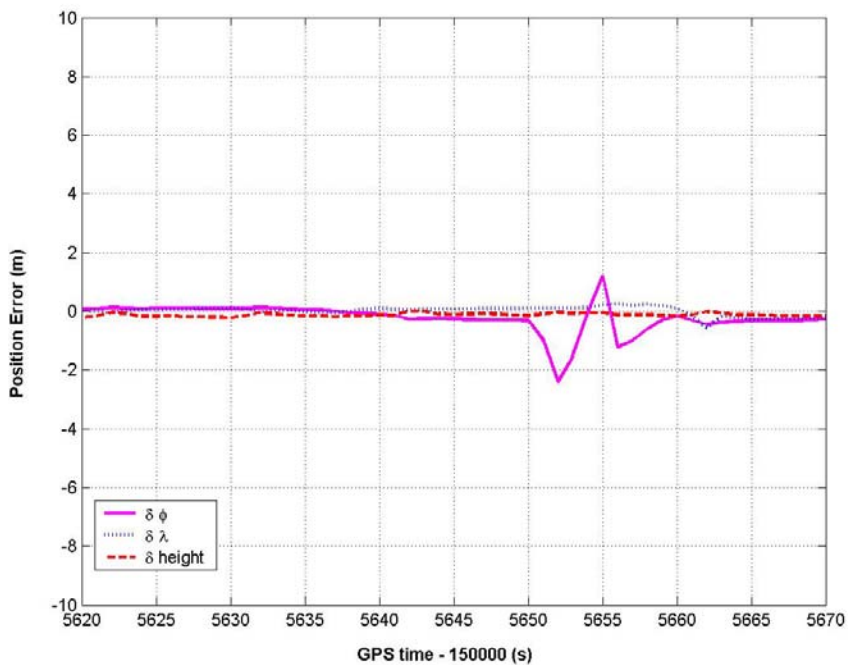


Figure 7.15b ANFIS-KF estimated position drift during outage (3)

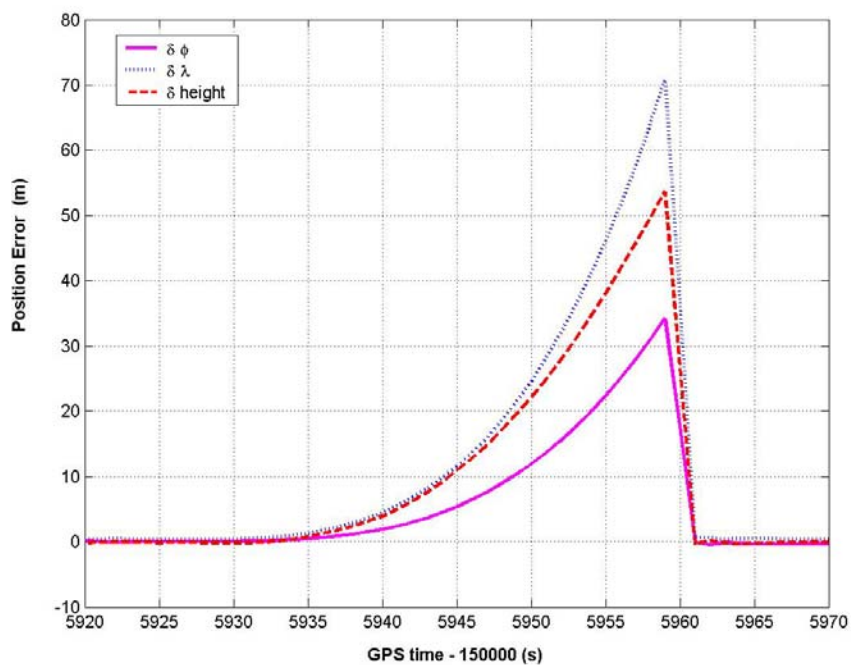


Figure 7.16a Kalman estimated position drift during outage (5)

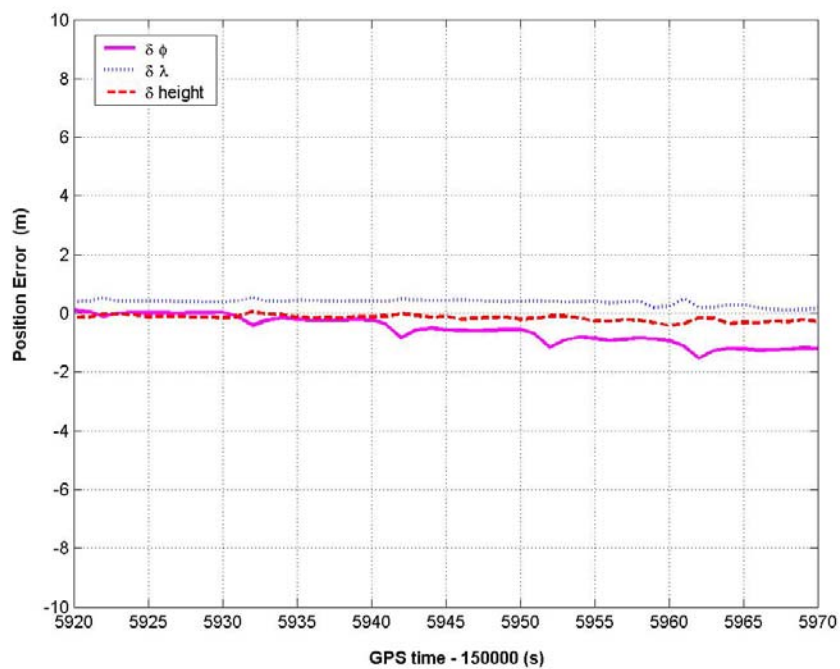


Figure 7.16b ANFIS-KF estimated position drift during outage (5)

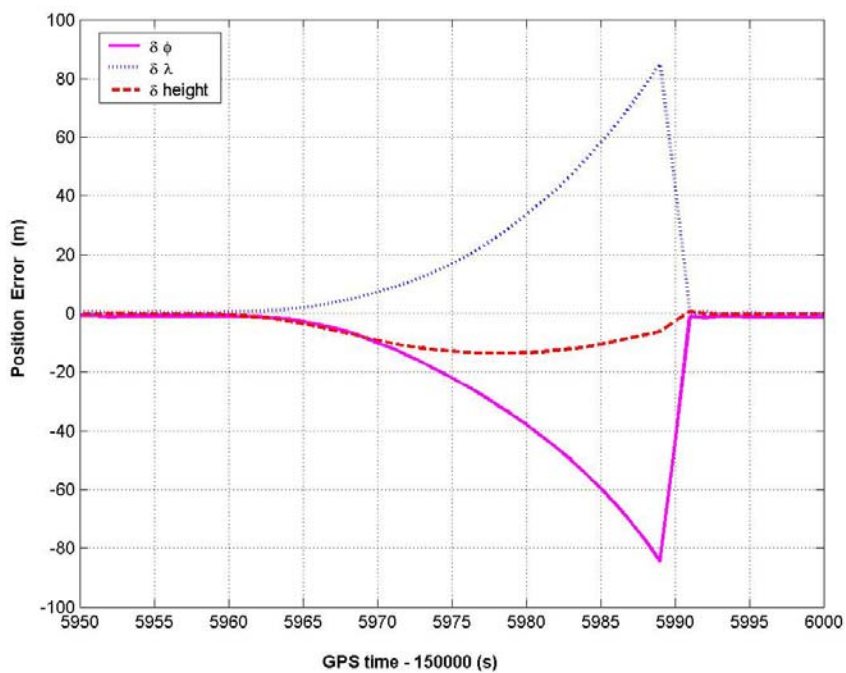


Figure 7.17a Kalman estimated position drift during outage (6)

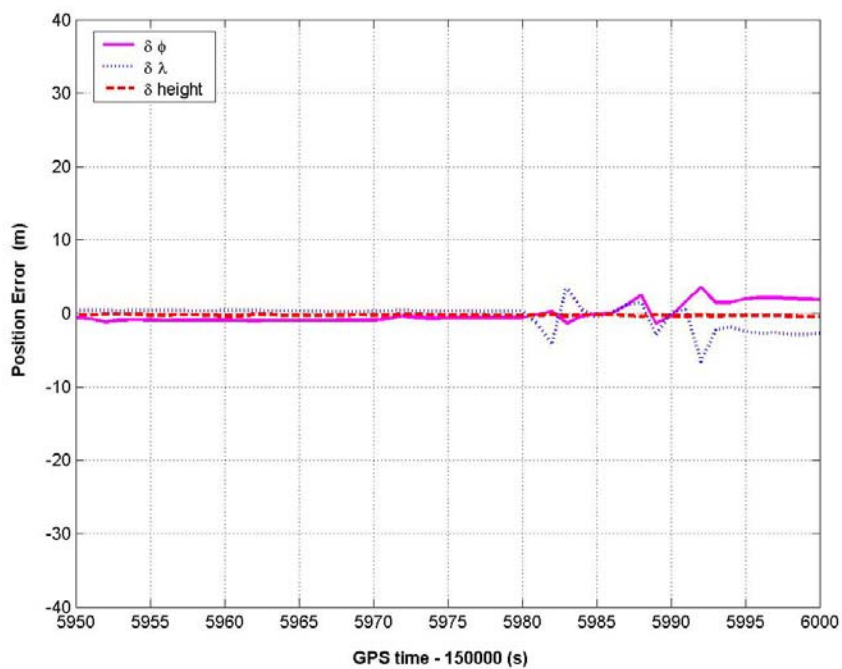


Figure 7.17b ANFIS-KF estimated position drift during outage (6)

Table 7.4 Maximum error values and RMSE of position and velocity during outage
(3)

	KF only		Proposed model	
	RMSE	Maximum Error	RMSE	Maximum Error
Err ϕ (m)	2.7	-1.7	0.7	-0.3
Err λ (m)	47.5	-118.9	0.1	0.2
Err h (m)	7.9	16.9	0.1	-0.2
Err V_E (m/s)	0.6	-1.7	0.6	1.1
Err V_N (m/s)	5.6	-11.3	0.3	-1.9
Err V_U (m/s)	0.7	-1.2	0.1	0.3

Table 7.5 Maximum error values and RMSE of position and velocity during outage
(5)

	KF only		Proposed model	
	RMSE	Maximum Error	RMSE	Maximum Error
Err ϕ (m)	13.8	34.2	0.6	-0.9
Err λ (m)	28.4	70.8	0.4	0.2
Err h (m)	23.1	53.7	0.2	-0.3
Err V_E (m/s)	1.6	3.3	0.1	-0.1
Err V_N (m/s)	3.3	7.1	0.1	-0.2
Err V_U (m/s)	2.3	-3.9	0.1	0.2

**Table 7.6 Maximum error values and RMSE of position and velocity during outage
(6)**

	KF only		Proposed model	
	RMSE	Maximum Error	RMSE	Maximum Error
Err ϕ (m)	13.3	-82.4	0.8	1.9
Err λ (m)	10.7	81.5	0.4	-2.6
Err h (m)	7.5	-17.6	0.2	-0.3
Err V_E (m/s)	1.7	-7.5	0.2	2.5
Err V_N (m/s)	1.5	7.7	0.5	4.6
Err V_U (m/s)	0.7	-1.8	0.2	0.2

Based on the previous results and analysis presented earlier in this section, some remarks and conclusions are in place. The proposed augmented ANFIS-KF method, for preserving the KF update functionality, has effectively reduced position and velocity drifts during the type of GPS outages simulated. Without using the proposed integrated Fuzzy-Kalman model, the long-term MEMS-INS standalone navigation accuracy of a traditional KF is unacceptable for many navigation applications, particularly land-vehicle navigation applications. Table 7.7 provides a brief comparison between the proposed Fuzzy-Kalman and the traditional KF solution listing the advantages and limitations of both techniques. Based on the results of the simulated case study and data set used in this

thesis, the proposed ANFIS-KF model, produced significant improvement of about 90% in the system's performance. As mentioned in Section 7.1.3, these results were obtained under the data set and the ideal simulated GPS signal outage scenarios used in this thesis. However, this can be considered as an initial proof of concept, validity, and reliability of the proposed ANFIS-KF augmentation.

Table 7.7 The proposed Fuzzy-Kalman versus the traditional KF

Comparison point	Proposed Fuzzy-Kalman	Traditional KF
Pre-knowledge of the system	Needs huge amounts of pre-stored input/output pattern data to develop the initial FIS structure	Needs a well defined model that fully describes the dynamics of the system
Appropriate System model	Modeless	Only works well under an appropriately predefined dynamic model
Short-term accuracy	Not guaranteed to provide better estimates than KF	Usually provides superior accuracy over short term (the first 5-10 sec)
Long-term accuracy	No accuracy degradation over time	Degradation of accuracy with the passage of time
Update measurements	No need for update measurements	Basic requirement to avoid accuracy degradation over time
Statistical Analysis	Not available since there is no exact error model to predict the error propagation	Errors can be easily propagated and covariance matrix of the estimates can be determined
Range of input data	Should be within the training input data ranges	Not restricted to any specific ranges of input data

CHAPTER EIGHT

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This Chapter summarizes the research work presented in this dissertation, draws conclusions from the dissertation's theoretical development and test results, and finally provides recommendations for future research and development in this field.

8.1 Summary

This research has two main objectives; First, to improve the performance of MEMS-based inertial sensors to facilitate GPS integration for land vehicle navigation applications; Second, to overcome the limitations and poor prediction of the conventional KF solution of MEMS-INS/GPS integration particularly during GPS signal or solution outage.

The WMRA is analysed and tested as a proposed tool to improve the performance of MEMS-based inertial sensors. A thorough analysis of MEMS-based accelerometer and gyroscope measurements in the frequency domain is carried out. The purpose of this primary step is to get an insight into the distribution of different types of inertial sensor errors prior to applying a wavelet transform. WMRA is then implemented to improve the sensors' signal to noise ratio, remove sensor errors mixed with motion dynamics, and provides more reliable data to the KF-based MEMS-INS/GPS integration module. Applying WMRA to MEMS inertial signals involves two main steps. The first step

involves eliminating the high frequency sensor noise using wavelet de-noising methods. The second step then follows by specifying a proper threshold level at which the motion dynamics can be separated from the short-term and/or long-term sensor errors as well as other disturbances. The proposed method has been verified using various data from a real navigation (van) test. The results have shown the effectiveness of the proposed WMRA method under the ideal testing scenarios used.

The variation of the MEMS sensors' performance characteristics with temperature and its effects on the accuracy of standalone inertial navigation parameters, particularly position, has been investigated. Intensive lab thermal testing of two typical MEMS-based IMUs (Crossbow AHRS400Ca and MotionPak II) was performed to study the trend and ranges of thermal variations for each individual sensor inside the two IMUs. The results have shown the non-linearity and repeatability of each sensor's thermal trend as well as the dissimilarity of the thermal trends of different sensors even those included in the same IMU. The ranges of thermal variations of gyros, for instance, were found to be relatively large (0.5 deg/sec) over the operating thermal range. Finally, the ANFIS-based networks prediction, known by its high capability of handling non-linear and unstable problems, has been suggested and assessed to model MEMS sensors' thermal behaviour. The focus of thermal modeling was mainly on gyro drift since it usually dominates the attainable accuracy of INS standalone navigation. Test results have shown the efficiency and precision of the proposed ANFIS modeling approach in estimating the thermal variation of gyro drift. Based on the obtained range of gyros thermal variations from lab tests, the

expected enhancement in positioning accuracy after applying ANFIS corrections may be very significant, based on the initial results obtained under specific conditions.

Two new techniques augmenting KF and Fuzzy principles for improving the integrated MEMS-INS/GPS system performance were presented. The first proposed technique provides direct corrections to the estimated position by Kalman during standalone inertial navigation mode (GPS reference solution not available in this case). The second proposed technique supplies estimated reference position and velocity error measurements to the KF update engine, therefore keeping its functionality during GPS signal outages. In both proposed techniques, an ANFIS model is extracted from input/output patterns. The input to the fuzzy system are the raw measurements of MEMS-based accelerometers and gyros involved in the IMU and the attitude angles determined through mechanization as well as GPS outage time. The prospective outputs of the fuzzy model, during GPS outages, are either position errors only, for the first proposed technique, or position and velocity errors together for the second proposed technique. As long as the GPS signals are available, the difference between the GPS and INS derived positions (or positions and velocity) are used as the desired output for online adaptation of the FIS parameters. Initial compacted FIS structure with its initial parameters is needed for ANFIS training. This initial FIS can be extracted first offline by applying the principles of Fuzzy-OLS method and by using intensive input/output data patterns covering different dynamics and motion scenarios. Both proposed techniques have been experimentally verified using limited MEMS inertial data collected in a land-vehicle navigation test. The experimental validation results have shown the efficiency and significant effect of the two proposed ANFIS-KF

augmented techniques in reducing position and velocity drifts during GPS outages under the scenarios tested.

8.2 Conclusions and Recommendations

The following conclusions can be drawn with respect to the methods developed and analyzed in this research, namely WMRA for de-noising MEMS inertial signals, ANFIS modeling for the thermal variations of MEMS inertial signals, as well as the augmented ANFIS-KF for solving MEMS-INS/GPS integrated system.

Wavelet-Multi-Resolution-Analysis

1. Through frequency domain analysis, the MEMS inertial output errors can be classified into long-term and short-term errors, which are primarily caused by the sensors' noise and secondarily by the dynamics resulting from road irregularities and engine vibrations.
2. Most long-term errors are reduced by updating the filter with the error state vector estimated in the GPS filter (position and velocity).
3. The short-term errors are partially reduced by the smoothing that results from the numerical integration process of the INS mechanization. The remaining component of errors lies in the frequency range 1-3 Hz, coinciding with the actual motion frequencies, which make it very difficult and risky to filter out these errors through the traditionally used low-pass-filters.

4. WMRA has shown its efficiency in the resolution of the noisy MEMS inertial output signal through its powerful multi-level decomposition approach, which facilitates the successful and safe separation of short-term errors from the true motion of interest.
5. WMRA can serve as a pre-filtering stage before processing the IMU data by the INS mechanization equations and the KF INS/GPS data fusion algorithm. This approach may improve the performance of both KF and mechanization processes by providing cleaner and more reliable data.
6. The results of experimental validation, using real inertial data, indicated that significant improvement of at least 50% in the positioning accuracy can be obtained after applying the proposed WMRA procedure.

ANFIS Modeling of Thermal Variation

1. The sensitivity of MEMS inertial sensors outputs to the variation of surrounding environmental conditions, particularly temperature, is relatively high due to their miniature design and fabrication process.
2. The repeatability of thermal variation of MEMS-based accelerometer biases or gyro drifts throughout the whole performed tests, indicates that a large part of the thermal effect is deterministic and can be modeled. However, the non-linearity and instability of this behaviour makes it difficult to model through traditional mathematical models like polynomials with their various orders.

3. The gyro thermal variations, for instance, were found to be relatively large (0.5 deg/sec) over the operating thermal range of the tests (-30 to 70 °C). Such an error, if not modeled and compensated for, may accumulate position errors, for instance, of hundreds - thousand of metres over short-term standalone MEMS-based inertial navigations (e.g. over 30 to 60 seconds)
4. The thermal behaviour of each individual MEMS-based inertial sensor (accelerometer or gyroscope) is specific and completely different from other sensors even for those of the same type and included inside the same IMU.
5. The above expected values of positioning errors in (3) necessitate the development of an accurate and reliable thermal modeling algorithm to compensate such thermal error particularly in real time navigation applications.
6. Following a thorough assessment of the proposed ANFIS modeling technique, its efficiency and precision in estimating thermal variation of gyro drift have been proven. After applying the thermal corrections through the proposed ANFIS-based modeling, the thermal error of gyro drift can be reduced to one hundredth (1/100) of its original value. This means reducing the accumulated positioning error by hundreds of metres over short-term standalone INS navigation. It is recognized however that the thermal calibration procedure would be costly to systematically implement in an operational environment.

ANFIS-KF Augmented Technique for MEMS-INS/GPS integration

1. The proposed augmented ANFIS-KF model provides more reliable position error estimates that can be used to correct the inaccurate estimates predicted by the KF during the absence of the DGPS position measurements.
2. The proposed model can effectively compensate for the absence of the update DGPS measurement and keep consistent positioning accuracy. In other words, the proposed augmented ANFIS-KF algorithm can appropriately bridge short-term GPS outages.
3. Without the use of the proposed model, the solution of KF is unacceptable for many applications when using MEMS-IMU and the system has a degraded performance during GPS outages.
4. The experimental validation of the first proposed ANFIS-KF augmentation approach for direct correction of estimated KF position has shown significant improvements under the limited scenarios tested.
5. Applying the second proposed ANFIS-KF augmentation approach for providing simulated reference measurements, exhibited significant improvement in the system performance, which is even better than the performance of the first proposed approach. However it can be said that both augmenting approaches have efficiently bridged the short-term (30 second) outage of reference GPS solution.
6. The Fuzzy-OLS algorithm is highly effective in reducing or compacting the size of the initial FIS structure, used for online ANFIS adaptive learning. This algorithm

reduces the number of antecedent and consequent parameters through the utilization of the interclass separability criterion as well as the orthogonal least-squares principles. This will definitely help decreasing the training time required by ANFIS to reach the convergence and therefore establish the TS-FIS model. As a result, more value is added to the proposed model especially in real time navigation missions where the time is always a critical issue.

In addition to the above listed conclusions, the following recommendations can be extracted,

1. A powerful pre-filter or signal analysing tool (like WMRA) other than the traditional low-pass and high-pass filters should be implemented to efficiently decompose the MEMS inertial sensors' output and identify the different comprised frequency components.
2. A thorough and comprehensive time and frequency domain pre-analysis of each sensor's output signal, in both static and dynamic modes, is essential. This will help select the proper denoising threshold in order to separate the true motion signal from other unwanted interfering signals, particularly those coinciding in frequency ranges with the motion of interest.
3. Modeling or correcting the thermal variation of MEMS-based inertial sensors output is vital and should not be neglected due to the significant impact of temperature changes on the miniature design and fabrication materials of these sensors.

4. A specific ANFIS thermal model should be developed separately for each individual sensor due to the dissimilarity of the different sensors behaviour even for those of the same type and included inside the same IMU.
5. Augmenting an efficient predictor (like ANFIS-based networks) with the traditional KF is crucial in case of integrating MEMS-based IMUs with GPS for navigation applications. This predictor should be capable of handling highly non-linear and imprecise systems in order to provide reliable state estimates rather than the poor ones produced by KF prediction engine in standalone inertial mode.
6. The test scenarios used in this thesis are idealistic; Excellent DGPS positions are available until the very moment total GPS outages are simulated. In reality, their will likely be a gradual geometry degradation and high noise and multipath affecting the GPS measurements prior to total loss of GPS position. As the geometry worsen, a criteria will have to be used to define when both the GPS measurements and geometry are too poor to be used. This will likely have a massive impact on the IMU performance in the following outage. Tests in actual degraded environments such as urban canyons have to be conducted to fully assess the above.

REFERENCES

Abdelazim, T.; Abdel-Hamid, W.; El-Sheimy, N. and Shin, E.H. (2004), "Experimental Results of an Adaptive Fuzzy Network Kalman Filtering Integration for Low Cost Navigation Applications", proceedings of NAFIPS-04 (North American Fuzzy Information Processing Society conference), Banff, Alberta, Canada

Abdel-Hamid, W. (2004), "AN ANFIS-based Modeling of Thermal Drift of MEMS-based Inertial Sensors", Proceedings of the ION-GNSS 2004 conference, Long Beach, California, USA, 21-24 Sep.

Abdel-Hamid, W.; Noureldin, A.; El-Sheimy, N. and Lachapelle, G. (2002), "Performance Analysis of MEMS-based Inertial Sensors for Positioning Applications", International Workshop System-on-Chip conference, Banff, 6-7 July 2002

Abdel-Hamid, W.; Osman, A.; El-Sheimy, N. and Noureldin, A. (2004), "Improving the Performance of MEMS-Based Inertial Sensors by Removing Short-Term Errors Utilizing Wavelet Multi-Resolution Analysis", Proceeding of the ION National Technical Meeting, California, USA, Jan. 2004

Abonyi J, Roubos JA, Oosterom M, and Szeifert F (2001) - Compact TS-Fuzzy Models through Clustering and OLS plus FIS Model Reduction - The 10th IEEE International Conference on Fuzzy Systems, Vol.(3), pp.1420 - 1423.

Abonyi, J.; Szeifert, F. and Babuska, R. (2001a), "Modified Gath-Geva fuzzy clustering for identification of takagi-sugeno fuzzy models," Submitted, 2001.

Alban, S. (2002), "An Inexpensive and Robust GPS/INS System for Automobiles", Proceedings of the ION-GPS 2002, Portland, OR, September 2002.

Aly, M.S.; Abdel-Hamid W. and El-Sheimy, N. (2004), "Noise Characterization of Micro-Machined Inertial Sensors", International Navigation Conference (MELAHA), Cairo, Egypt, 13-15 Apr. 2004

Andrews, T.R. (1975), "Standard gyro and accelerometer terminology", Compiled by Gyro and Accelerometer Panel, Aerospace and Electronic Systems Society, IEEE.

Astrom K. J. and Wittenmark B. (1984), "Computer Controller Systems: Theory and Design", Prentice-Hall, Inc., 1984

- Babuska, R. (1998), "Fuzzy Modeling for Control", Kluwer Academic Publishers, Boston, 1998.
- Babuska, R. and Verbruggen, H.B. (1997), "Fuzzy set methods for local modeling and identification, Multiple Model Approaches to Non-linear Modeling and Control", R. Murray-Smith and T. A. Johansen, Eds., pp. 75–100. Taylor & Francis, London, UK, 1997.
- Bar-Itzhack, I.Y. and Berman, N. (1988), "Control theoretic approach to inertial navigation system", AIAA Journal of Guidance, Control & Dynamics, Vol.(11), pp. 237-245.
- Bauersima, I. (1983), "NAVSTAR/Global Positioning System (GPS) II", Radiointerferometrische Satellitenbeobachtungen, Mitteilungen der Satelliten-Beobachtungsstation Zimmerwald, Bern, Vol. 10.
- Bernstein J. et al. (1998), "Low Noise MEMS Vibration Sensor for Geophysical Applications," Proc 1998 Solid State Sensor and Actuator Workshop, Hilton Head Island, SC:55-58. Extended version also in IEEE JMEMS Dec. 1999.
- Bernstein, J. and Weinberg, M.S. (1996), "Comb-Drive Micromechanical Tuning Fork Gyro Fabrication Method", U.S. Patent #5,496,436.
- Bernstein, J. et al. (1993), "A micromachined comb-drive tuning fork rate gyroscope," Proc IEEE Micro Electro Mechanical Systems Workshop (MEMS '93), Fort Lauderdale, pp 143-148.
- Bertolini, A. (2000), "High sensitivity accelerometers for gravity experiments", PhD thesis, California Institute of Technology, 2000.
- Billings, S.A.; Korenberg, M.J.; and Chen, S. (1988), "Identification of non-linear output affine systems using an orthogonal least-squares algorithm", International Journal of Systems and Science, Vol. (19), pp.1559-1568.
- Birlasekaran, S. and Ming, Y. (2002), "Characterization of partial discharge signals using wavelet and statistical techniques", IEEE International Symposium on Electrical Insulation, pp. 9-13.
- Bishop, C.M. (1995), "Neural Networks for Pattern Recognition", Oxford University Press, 1995.

Braasch, M. (1998), "GPS Multipath Characterization and Mitigation", Lecture Notes, Geomatics Engineering Department, The University of Calgary, July 8.

Bracewell, R. N. (1965), "The Fourier Transform and Its Applications", New York: McGraw-Hill Book Company, 381.

Brault, J. W. and White, O. R. (1971), "The analysis and restoration of astronomical data via the fast Fourier transform", *Astron. & Astrophys.*, 13, pp. 169-189.

Brenner, M. (1996), "Integrated GPS/Inertial Fault Detection Availability", *Navigation*, Vol 43, No. 2, 1996, pp. 339-358.

Brigham, E. O. (1988), "The Fast Fourier Transform and Its Applications", Englewood Cliffs, NJ: Prentice-Hall, Inc., 448.

Brown, R. G. and Hwang, P. C. (1985), "Introduction to random signals and applied Kalman filtering", Second edition, John Wiley & Sons Inc., 1985

Bruton, A.M.; Skaloud, J. and Schwarz, K.P. (1999), "The Use of Wavelets for The Analysis and De-noising of Kinematic Geodetic Measurements", *Geodesy Beyond 2000-IAG General Assembly*, Birmingham, UK, July 19-24.

Burrus, C.S.; Gopinath, R. and Guo, H. (1998), "Introduction to wavelet and wavelet transforms a primer", Prentice Hall, 1998.

Chiang, K. W. (2004), "INS/GPS integration using Neural Networks for navigation applications", PhD Thesis, Department of Geomatics Engineering, University of Calgary, November 2004.

Chiang, K.W.; Noureldin, A. and El-Sheimy, N. (2003), "Multi-sensors Integration using Neuron Computing for Land Vehicle Navigation", *GPS Solutions*, Springer-Verlag Heidelberg, V6 (4), pp: 209 – 218, March 2003.

Chui, C.K. (1992), "An Introduction to Wavelets", Academic Press, San Diego, CA, 1992.

Cousins, F. W. (1990), "The anatomy of a the gyroscope", in Hollington, J.H. (Ed.): *Agardograph 313*, Parts I, II and III, (AGADR, 1990).

Daubechies, I. (1988), "Orthonormal basis of compactly supported wavelets", *Comm Pure Applied Math* 41, pp. 909-996.

Davis, B. S. (1998), "Using low cost MEMS accelerometer and Gyroscope as strap-down IMUs on rolling projectiles", IEEE, 1998

DeLoach, S.R. and Remondi, B. (1991), "Decimeter positioning for dredging and hydrographic surveying", Proceedings of the First International Symposium on Real Time Differential Applications of the Global Positioning System. TUV Rheinland, Koln, Vol. 1, pp 258-263.

Diesel, J. W. and Lu, S. (1995), "GPS/IRS AIME: Calculation of Thresholds and Protection Radius Using Chi-Square Methods", Proceedings of ION-NTM, Palm Springs, CA. Sep. 1995. pp 1959-1964.

Donoho, D.L. (1995), "De-noising by soft-thresholding", IEEE Transactions on Information Theory 41(3), pp. 613-627.

Donoho, D.L. and Johnstone, I.M. (1994), "Ideal spatial adaptation by wavelet shrinkage", Biometrika 81, pp. 425-455.

El-Rabbany, A. (1996), "Temporal characteristics of Multipath Errors", Technical Report No. 1044, Royal Institute of Technology, Department of Geodesy and Photogrammetry, Stockholm, Sweden.

El-Sheimy, N. (2002) "Introduction to Inertial Navigation", ENGO 699.71 lecture notes, Department of Geomatics Engineering, the University of Calgary.

El-Sheimy, N. (2004), "Inertial Surveying and INS/GPS Integration", ENGO 623 lecture notes, Geomatics department, University of Calgary.

Fahlman S. E. (1988), "Faster-learning variations on back-propagation: an optimal study", Proceedings of the 1988 Connectionist Models Summer School, pages 38-51, Carnegie Model University.

Farrell, J. and Barth, M. (1999), "The Global Positioning System and Inertial Navigation", McGraw-Hill, New York, NY, 1999

Fedder, G.K. (1994), "Simulation of Micro-electro-mechanical Systems", PhD thesis, UC Berkeley, 1994.

Gabrielson, T.B. (1993), "Mechanical-thermal noise in micromachined acoustic and vibration sensors," IEEE Trans, Electron. Devices, Vol. 40:903-909.

Geen, J. (1997), "Micromachined device with rotationally vibrated masses", U.S. Patent #5,635,640.

Gelb, A. (1989), "Applied optimal estimation", Text Book, The Analytic Science Corporation, 1989

Gershenfeld, N.; Schoner, B. and Metois, E. (1999), "Cluster-weighted modelling for time-series analysis," Nature, vol. 397, pp. 329–332, 1999.

Harris, C.M. and Crede, C.F. [Eds.] (1961), "Shock and vibration hand book", McGraw-Hill, 1-3

Helsel, M.; Gassner, G.; Robinson, M. and Woodruff, J. (1994), "A Navigation Grade Micro-Machined Silicon Accelerometer", Proceedings of IEEE, Vol 2, 1994, pp 51-58.

Hide, C.; Moore, T. and Smith, M. (2002), "Adaptive Kalman Filtering for Low Cost INS/GPS", ION GPS 2002, Portland, OR, 24-27 September 2002.

Hoffmann-Wellenhof, B.; Lichtenegger, H. and Collins, J. (1997), "Global Positioning System Theory and Practice", Springer-Verlag Wien New York.

Hou, H. and El-Sheimy, N. (2003), "Inertial Sensors Errors Modeling Using Allan Variance", Accepted for publication, ION GPS/GNSS 2003 Proceedings, Sep 9-12, Portland, OR 2003. . (Winner of ION best paper award for Inertial Navigation Systems)

Huiping, H.; Binzhen, Z. and Wendong, Z. (2001), "Study of accelerometer's temperature compensation by software and hardware", Proceedings of the International Symposium on Test and Measurement, v 1, 4th International Symposium on Test and Measurement (ISTM/2001), Jun 1-3 2001, Shanghai, pp 841-843.

Hulsing, R. (1998), "MEMS inertial rate and accelerometer sensor", IEEE Aerospace and Electronic Systems Magazine, Vol. 13, pp (17-23), 1998

IEEE std 1139 (1988), "IEEE Standard Definitions of physical quantities for Fundamental Frequency and Time Metrology".

IEEE std 1293 (1998), "IEEE Standard Specification Format Guid and Test Procedure for Linear, Single-Axis, Non-gyroscopic Accelerometers".

IEEE std 952 (1997), "IEEE Standard Specification Format Guido and Test Procedure for Single-Axis Interferometric Fiber Optic Gyro".

Jang, J. (1991), "Fuzzy modeling using generalized neural networks and Kalman filter algorithm", Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91), pp 762-767, July 1991.

Jang, J. (1993), "Self-learning fuzzy controllers based on temporal back propagation", IEEE Transactions on Neural Networks, Vol. 3 No. 5, pp. 714-23

Jang, J. (1996), "Input selection for anfis learning," Proceedings of the IEEE International Conference on Fuzzy Systems. New Orleans, 1996.

Jang, J.; Sun, C. and Mizutani, E. (1997), "Neuro-Fuzzy and soft computing, A computational approach to learning and machine intelligence", Prentice Hall Inc., Englewood Cliffs, N.J., USA.

Jay, A.F. and Matthew, B. (1999), "The Global Positioning System & Inertial Navigation", Mc-Graw Hill NY.

KINGSPAD (2000), "Documentation for KINGSPAD - KINematic Geodetic Software for Position and Attitude Determination - Version 2.1", Department of Geomatics Engineering, The University of Calgary, Canada.

Kirkpatrick, E.G. (1974), "Introductory statistics and probability for engineering science and technology", Prentice-Hall.

Klobuchar, J. (1991), "Ionospheric Effects on GPS", GPS World, Volume 2, Number 4.

Lachapelle, G. (2004), "Navstar GPS: Theory and Applications", ENGO 625 lecture notes, Fall 2004, Department of Geomatics Engineering, The University of Calgary, Canada.

Lachapelle, G.; Cannon, M.E. and Lu, G. (1992), "High-precision GPS Navigation with Emphasis on Carrier Phase Ambiguity Resolution", Marine Geodesy, Vol. 15.

Lachapelle, G.; Falkenberg, W. and Casey, M. (1987), "Use of Phase Data for Accurate GPS Differential Positioning", Proceedings of PLANS, Las Vegas, Nov. 5-7.

Langley, R. (1993), "The GPS Observables", GPS World, Volume 4, Number 4.

Lee, C.C. (1990), "Fuzzy logic in control systems: Fuzzy logic controller", IEEE Transactions on Systems, Man and Cybernetics, Vol.(20), pp.404-418.

Lin, J. W.; Hwang, M. I.; and Becker, J. D. (2003), "A fuzzy neural network for assessing the risk of fraudulent financial reporting", *Managerial Auditing Journal* 18/8 [2003] pp 657-665.

Lingjun, M.; Yonghong, L. and Wendong, Z. (2001), "Study on MEMS accelerometers' 0g offset temperature drift compensation", *Proceedings of the International Symposium on Test and Measurement*, v 1, 4th International Symposium on Test and Measurement (ISTM/2001), Jun 1-3 2001, Shanghai, p 774-776.

Lutz, M. et al. (1997), "A precision yaw rate sensor in silicon micromachining," *Tech Dig 9th Intl. Conf Solid State Sensors and Actuators (Transducers '97)*, Chicago, IL:847-850.

Malhotra, R. and Malhotra, D. (1999), "Fuzzy systems and neuro-computing in credit approval", *Journal of Lending & Credit Risk Management*, Vol. 81 No. 11, pp. 24-7.

Mallat, S. (1989), "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Trans Patt Recog and Mach Intell* 11(7), pp. 674-693.

Mamdani, M. (1974), "Application of fuzzy algorithm for control of simple dynamic plant", *Proc. IEE*, Vol.(121), pp.1585-1588.

Mamdani, E. and Assilian, S. (1975), "An experiment in linguistic synthesis with a fuzzy logic controller", *International Journal of Man-machine Studies*, Vol. 7 No. 1, pp. 1-13.

Mathwork's Help Documents (2002), "User Guide for Fuzzy Logic Tool Box", Mathwork Inc., Release 13, version 2.1.2, 9 July, Natick, MA, USA

Merhav, S. (1996), "Aerospace sensor systems and applications", Springer-Verlag, New York, Inc.

Mohamed, A.H. (1999), "Optimizing the Estimation Procedure in INS/GPS Integration for Kinematic Applications, [PhD thesis]", UCGE Report No. 20127, Department of Geomatics Engineering, The University of Calgary, Canada.

Mohinder S.G. and Angus P.A. (2001), "Kalman Filtering: Theory and Practice Using MATLAB", 2nd Edition, ISBN: 0-471-39254-5, Wiley InterScience.

Murray-Smith, R. and Johansen, T.A. (1997), "Multiple Model Approaches to Nonlinear Modeling and Control", Taylor & Francis, London, UK, 1997.

Nassar, S. (2003), "Improving the Inertial Navigation System (INS) Error Model for INS and INS/DGPS Applications", PhD thesis, UCGE Reports Number 20183, Geomatics department, University of Calgary, Canada

Nebot, E. and Durrant-Whyte, H. (1999), "Initial calibration and Alignment of low cost inertial navigation units for land vehicle applications", *Journal of Robotic Systems*, Vol. 16, pp (81-92), Feb. 1999

Ogden R.T. (1997), "Essential wavelets for statistical applications and data analysis", Birkhäuser Boston, 1997.

Parkinson, B.W. and Spilker, J.J. (editors) (1996), "Global Positioning System: Theory and Applications", American Institute of Aeronautics and Astronautics, Inc.

Phillips, R.E. and Schmidt, G.T. (1996), "GPS/INS Integration", AGARD lecture series 207, pp 9-1 to 9-18, 1996

Pimbley, J. M. (1992), "Recursive Autoregressive Spectral Estimation by Minimization of the free Energy", *Proceedings of The IEEE Transactions on Signal Processing*, V.40, No.6, 1518-1527, June 1992.

Pottenger, M.D. and Kaiser, W.J. (2001), "Modeling and simulation of micro-systems", www.cr.org, ISBN 0-9708275-0-4.

Putty, M.W. (1995), "A micromachined vibrating ring gyroscope," Ph.D. dissertation, University of Michigan, Ann Arbor, MI.

Reda-Taha, M. M.; Noureldin, A. and El-Sheimy, N. (2003), "Improving INS/GPS Positioning Accuracy During GPS Outages Using Fuzzy Logic", *Proceedings of the 16th annual Technical Meeting of Satellite Division, Institute of Navigation (ION) GPS-GNSS 2003*, Portland, Oregon, Sept. 2003, pp. 499-508.

Remtema T. and Lin, T. (2000), "Active Frequency Tuning for Micro-resonators by Localized Thermal Stressing Effects," *Solid-State Sensor and Actuator Workshop*, pp. 363–366, (Hilton Head Island, SC), June 2000.

Robertson, D.C.; Camps, O.I.; Mayer, J. S. and Gish, W.B. (1996), "Wavelets and electromagnetic power system transients", *IEEE Trans. on Power Delivery*, Vol. 11, No. 2, April 1996, pp. 1050-1056.

Roubos, J.A.; Setnes, M. and Abonyi, J. (2001), "Learning fuzzy classification rules from data," *Developments in Soft Computing*, R. John and R. Birkenhead, Eds. 2001, Springer, Physica Verlag.

Salychev, O. S.; Voronov, V. V.; Canon, M. E.; Nayak, R. and Lachapelle, G. (2000), "Low Cost INS/GPS Integration: Concepts and Testing", Institute of Navigation National Technical Meeting/Anaheim, CA/January 26-28, 2000

Salychev, O.S. (1998), "Inertial Systems in Navigation and Geophysics", Moscow State Technical University, Bauman MSTU Press, Moscow.

Savet, P.H. (1961), "Gyroscopes: theory and design", McGraw Hill, 1961

Schwarz, K.P. (1983), "Inertial Surveying and Geodesy", *Review of Geophysics and Space Physics*, Vol. 21, No. 4, pp. 878-890.

Schwarz, K.P. (1986), "The Error Model of Inertial Geodesy – a Study in Dynamic System Analysis", in H. Sünkel (ed.) "Mathematical and Numerical Techniques in Physical Geodesy", Springer-Verlag, Berlin.

Schwarz, K.P. and El-Sheimy, N. (1999), "Future Positioning and Navigation Technologies", Final Report, Study performed at the University Calgary under Batelle, Columbus, USA, contract no. DAAH04-96-C-0086, 1999

Schwarz, K.P. and Wei, M. (1990), "Testing a Decentralized Filter for GPS/INS Integration", *Proceedings of the IEEE Position, Location and Navigation Symposium*, 1990, pp. 429-435

Schwarz, K.P. and Wei, M. (2000), "Inertial Geodesy and INS/GPS Integration", partial lecture notes for ENGO 623, Department of Geomatics Engineering, The University of Calgary, Canada.

Seeber, G. (1993), "Satellite Geodesy, Foundations, Methods, and Applications", Walter de Gruyter, Berlin, Germany.

Setnes, M. and Hellendoorn, H. (2000), "Orthogonal transforms for ordering and reduction of fuzzyrules," *FUZZ-IEEE*, 2000, pp. 700–705.

Setnes, M.; Babuska, R. and Verbruggen, H.B. (1998), "Rule-based modeling: Precision and transparency," *IEEE Trans. SMC-C*, vol. 28, pp. 165–169, 1998.

Shcheglov, K.; Evans, C.; Gutierrez, R. and Tang, T.K. (2000), "Temperature dependent characteristics of the JPL silicon MEMS gyroscope", IEEE Aerospace Conference Proceedings, v 1, Mar 18-25 2000, Big Sky, MT, p 403-411.

Shin, E.H. (2001), "Accuracy Improvement of Low Cost INS/GPS for Land Applications", MSc thesis, UCGE Reports Number 20156, Geomatics department, University of Calgary, Canada

Shin, E.H., and El-Sheimy, N. (2002), "INS Tool Box, INS/GPS integration software", Mobile multi-sensors research group, Department of Geomatics Engineering, the University of Calgary

Shkel, A.; Howe, R.T. and Horowitz, R. (1999), "Modeling and Simulation of Micro-machined Gyroscopes in the Presence of Imperfections," International Workshop on Micro Robots, Micro Machines and Systems, September/October 1999.

Skaloud J. (1999), "Optimizing Georeferencing of Airborne Survey Systems by INS/DGPS", Ph.D. Thesis, The University of Calgary, Dept. of Geomatics Engineering, Calgary, Alberta, 1999.

Skaloud J.; Bruton, A.M. and Schwarz, K.P. (1999), "Detection and Filtering of Short-Term (1/f) Noise in Inertial Sensors", NAVIGATION, Journal of The Institute of Navigation, Vol. 46, No. 2: 97-107, Summer 1999.

Smith, J. G. and Balleto, A. (1994), "Dynamic Admittance Matrix of piezoelectric cantilever bimorphs", Journal of Micro-electro-mechanical Systems, Vol. 3, No. 3, Sep. 1994

Smithson, T.G. (1987), "A review of the mechanical design and development of a high performance accelerometer", Proc. Inst. Mech. Eng., Conference on mechanical technology of inertial sensors, C49/87.

Song, C. (1997), "Commercial vision of silicon based inertial sensors," Tech Dig 9th Intl. Conf Solid State Sensors and Actuators.

Stephen, J. (2000), "Development of a Multi-Sensor GNSS Based Vehicle Navigation System", M.Sc. Thesis, UCGE Report No. 20140, Department of Geomatics Engineering, University of Calgary, August 2000

Strang, G. and Nguyen, T. (1997) "Wavelets and Filter Banks", Wellesley-Cambridge Press, 1997.

Sugeno, M. (1985), "Industrial Applications of Fuzzy Control", Elsevier Science, Amsterdam.

Takagi T, and Sugeno M (1985) - Fuzzy identification of systems and its application to modeling and control - IEEE Transactions on Systems, Man and Cybernetics, Vol.(15), pp.116–132.

Tang, T.K. et al. (1997), "A packaged silicon MEMS vibratory gyroscope for micro-spacecraft," Proc IEEE Micro Electro Mechanical Systems Workshop (MEMS '97), Japan:500-505.

Tehrani, M. M. (1983), "Ring Laser Gyro Data Analysis with Cluster Sampling technique," Proceeding of SPIE, vol 412, 1983.

Titterton D.H. and Weston, J.L. (1997): "Strapdown inertial navigation technology;" Peter Peregrinus Ltd., London, UK, 1997.

Tsoukalas, L. H. and Uhrig, R. E. (1997), "Fuzzy and neural approaches in engineering", 1st edition, Wiley, NY, USA.

Turban, E. and Aronson, J. (2000), "Decision Support Systems and Intelligent Systems", 6th ed., Prentice-Hall, Englewood Cliffs, NJ.

Turčan A, Ocelíková E, and Madarász L (2003) - Fuzzy C-means Algorithms in Remote Sensing - Proceedings of the 1st Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence (SAMI), Herlany, pp.207-216.

Varadan, V. K.; Varadan, V. V. and Subramanian, H. (2001), "Fabrication, characterization and testing of wireless MEMS-IDT based micro-accelerometer", Sensors and Actuators, A: physical, Vol. 90, pp (7-19), May 2001

Walker, J.S. (1999), "Wavelets and their scientific applications", Chapman & Hall/CRC, New York, 1999.

Wells, D.E.; Beck, N.; Delikaraoglou, D.; Kleusberg, A.; Krakiwsky, E.J.; Lachapelle, G.; Langley, R.g.; Nakibolou, M.; Schwarz, K.P.; Tranquilla, J.M. and Vanicek, P. (1986), "Guide to GPS Positioning", Canadian GPS Associates, Fredericton, New Brunswick.

Werbos, P. (1974), "Beyond regression: New tools for prediction and analysis in the behavioural sciences", PhD thesis, Harvard University, 1974

Wild, U.; Beutler, G.; Frankhauser, S. and Gurtner, W. (1990), "Stochastic Properties of the Ionosphere Estimated from GPS Observations", Proceedings of the International Symposium on Precise Positioning with the Global Positioning System, Ottawa, Canada, September 3-7.

Wong, R.V.C. (1988), "Development of a RLG Strapdown Inertial Navigation System, [PhD thesis]", USCE Report Number 20027, Dept. of Geomatics Eng., University of Calgary, Canada.

Yazdi, N.; Ayazi, F. and Najafi, K. (1998), "Micromachined Inertial Sensors," Proc IEEE, Vol. 86, No. 8.

Yeh, C. and Najafi, K. (1997), "A low voltage tunnelling-based silicon micro-accelerometer", IEEE Transaction on Electron Devices, Vol. 44, No. 11, Nov. 1997

Yen, J. and Wang, L. (1999), "Simplifying fuzzy rule-based models using orthogonal transformation methods," IEEE Trans. SMC-B, vol. 29, pp. 13-24, 1999.

Zadeh, L. (1984), "Making Computers Think Like People", IEEE Spectrum, Vol.(7), pp 26-32.

Zadeh, L. (1994), "Fuzzy logic, neural networks and soft computing", Communication of the ACM, Vol. 37 No. 3, pp. 77-84.

Zhou, Y. (1998), "Layout Synthesis of Accelerometers", M.Sc. Project Report, Department of Electrical and Computer Engineering, Carnegie Mellon University, August 1998